VOLUME 102 PART B NUMBER 4 JULY 1955

The Proceedings

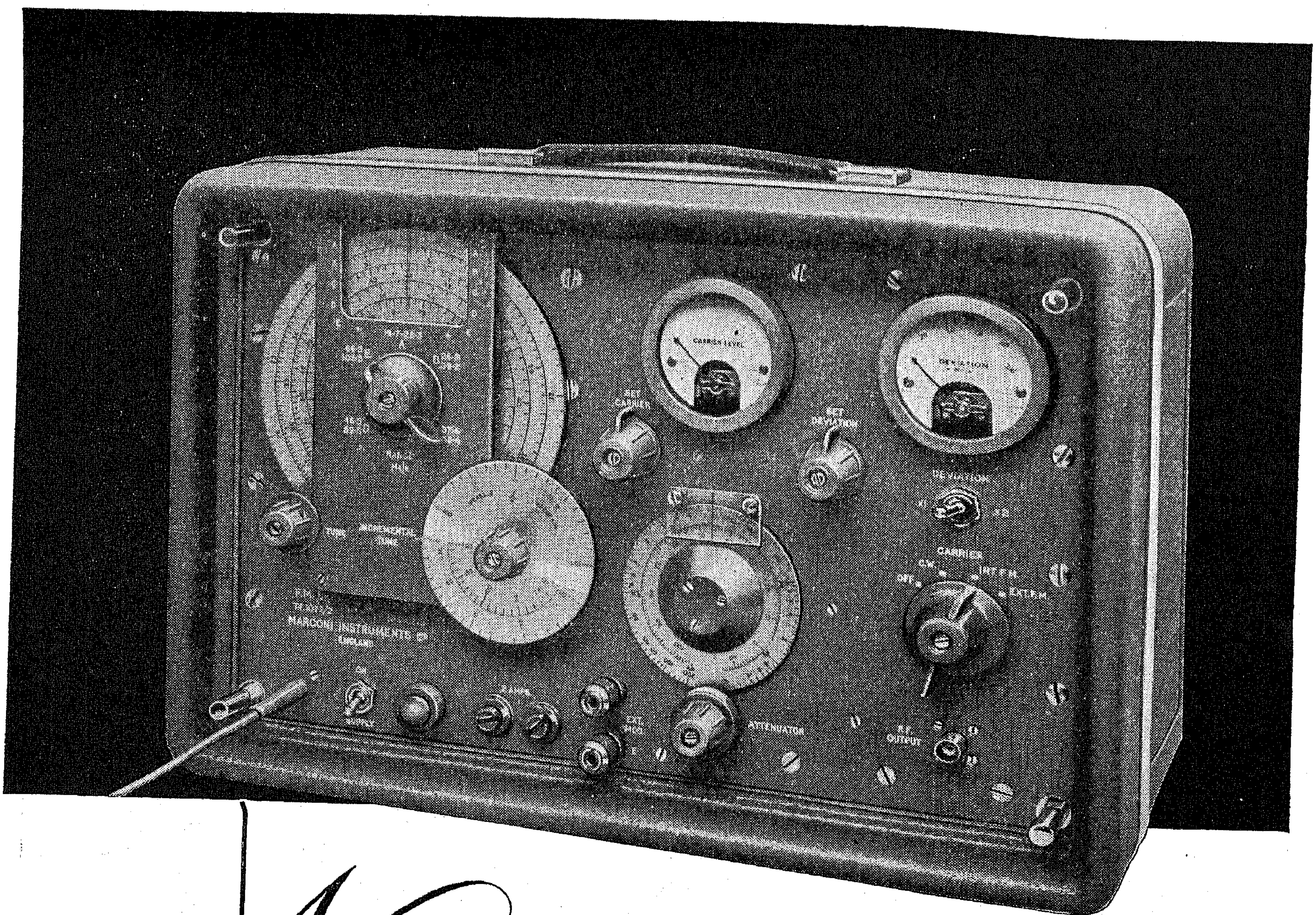
OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

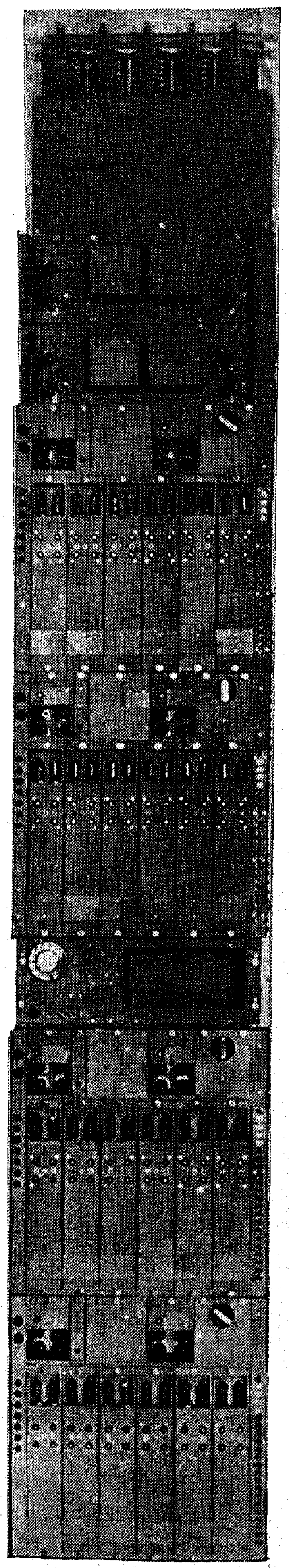


SKILLMAN

Twelve Channel Groups

COMPLETE WITH DIALLING EQUIPMENT

for use on



CARRIER TYPE CABLES

12, 24 or 48 Channel groups are available conforming in all respects to C.C.I.F. recommendations.

COAXIAL CABLES

60 channel Super Groups are available to C.C.I.F. recommendations.

RADIO LINKS

Up to 120 Channels can be provided on existing links or in conjunction with new Murphy developments (see opposite).

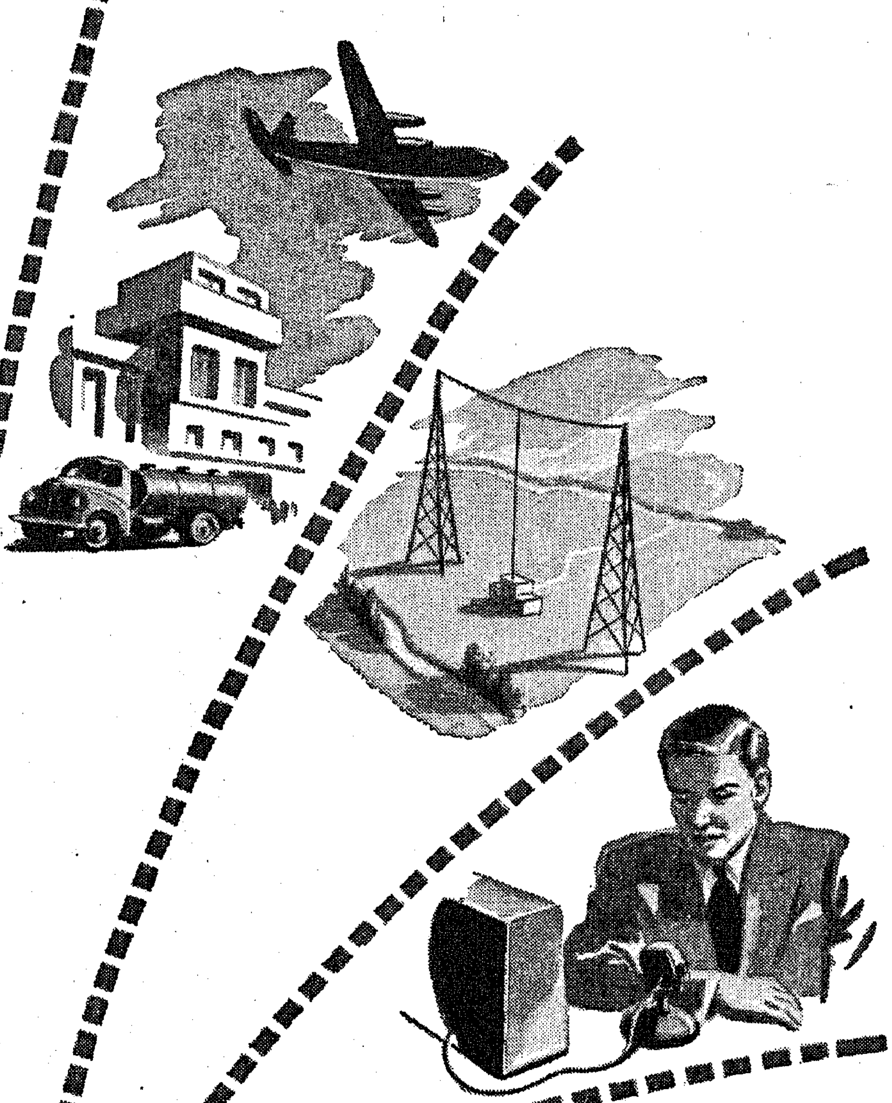
★ Each channel reaches the switching equipment as a voice circuit plus the equivalent of two extra copper wires from end to end for signalling.

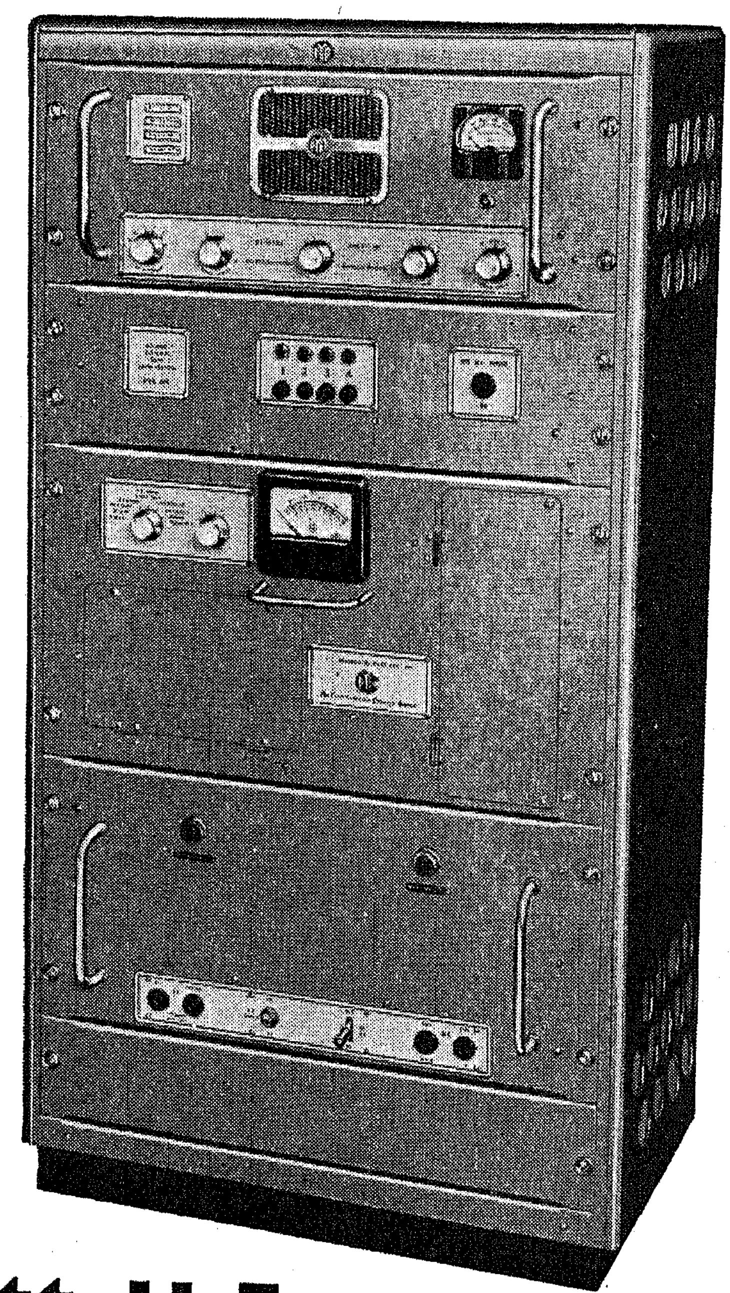
The advantages of this from standardisation and organisational points of view are becoming widely recognised and, in making any economic studies, the complete elimination of the large groups of relays, required for pulse signalling systems, should not be overlooked.

★ Can be combined with T.O.M.E.

OTHER SYSTEMS AVAILABLE FOR USE ON JUNCTION CABLES OR OPEN WIRE LINES

The illustration shows 24 channel ends, complete with signalling gear, 2 wire V.F. terminations, amplifiers, etc., all mounting on one side of a rack: 48 channels will mount on a double-sided rack.



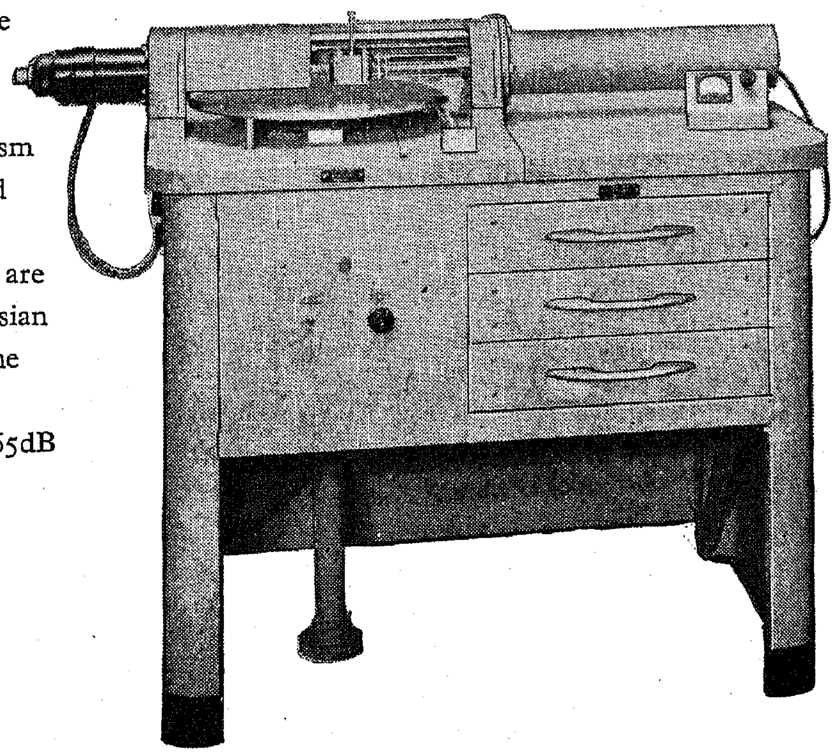


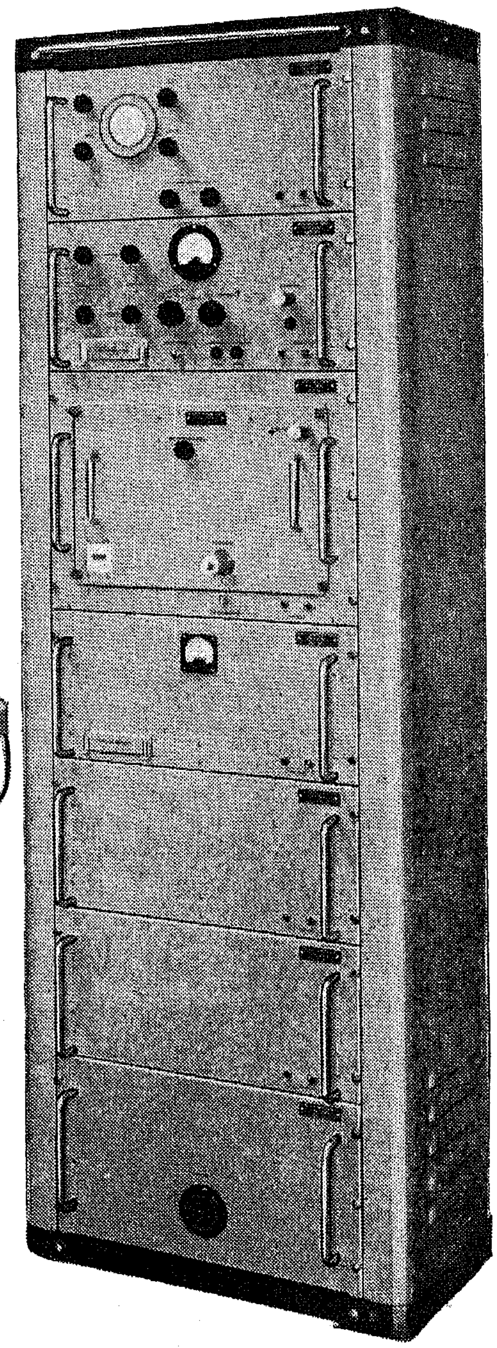
AUTOMATIC ANTENNA PATTERN RECORDING

EKCO ANTENNA PATTERN RECORDER type E59

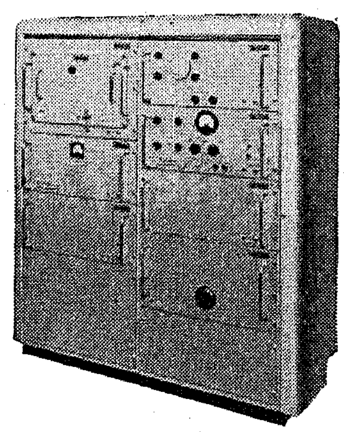
This equipment has been developed by EKCO Electronics to automatically record the radiation patterns of any centimetric antenna. The antenna under test is mounted on the roof of a rotatable trailer and illuminated by a fixed transmitter. The amplitude of the received signal is then continuously plotted against the angular traverse of the trailer.

All the equipment except the transmitter unit is mounted in the trailer and remote controls for the transmitter are provided. The received C.W. signal is mixed with a modulated local oscillator signal and the resultant I.F. output combined with an anti-phase modulated I.F. signal. The reference signal is derived from a 30 Mc/s oscillator and servo-driven piston attenuator. The combined signals are fed via a seven-stage, low-noise I.F. amplifier to a balanced modulator, and the resultant error signal applied to a servo amplifier. The output of this amplifier drives a servo motor which moves the piston attenuator in such a direction as to reduce the difference between the reference signal and the received signal. A pen attached to the piston drive mechanism records the amplitude of the received signal in terms of the attenuation law of the standard piston. Facilities are available for plotting either on Cartesian or polar co-ordinate graph paper. The amplitude scale in each case is 10dB per inch with a maximum travel of 65dB and the Cartesian co-ordinate paper can be run at rates corresponding to 2 or 5 degrees per inch.





The Standard Equipment covers the frequency range of 9300 to 9500 m/c. The frequency changer section is a plug-in unit and other frequency ranges can be covered to special order. The Recorder can be supplied in two forms with either a single 6ft. (shown above with plotting table) or a twin 4 ft. console rack shown at right.

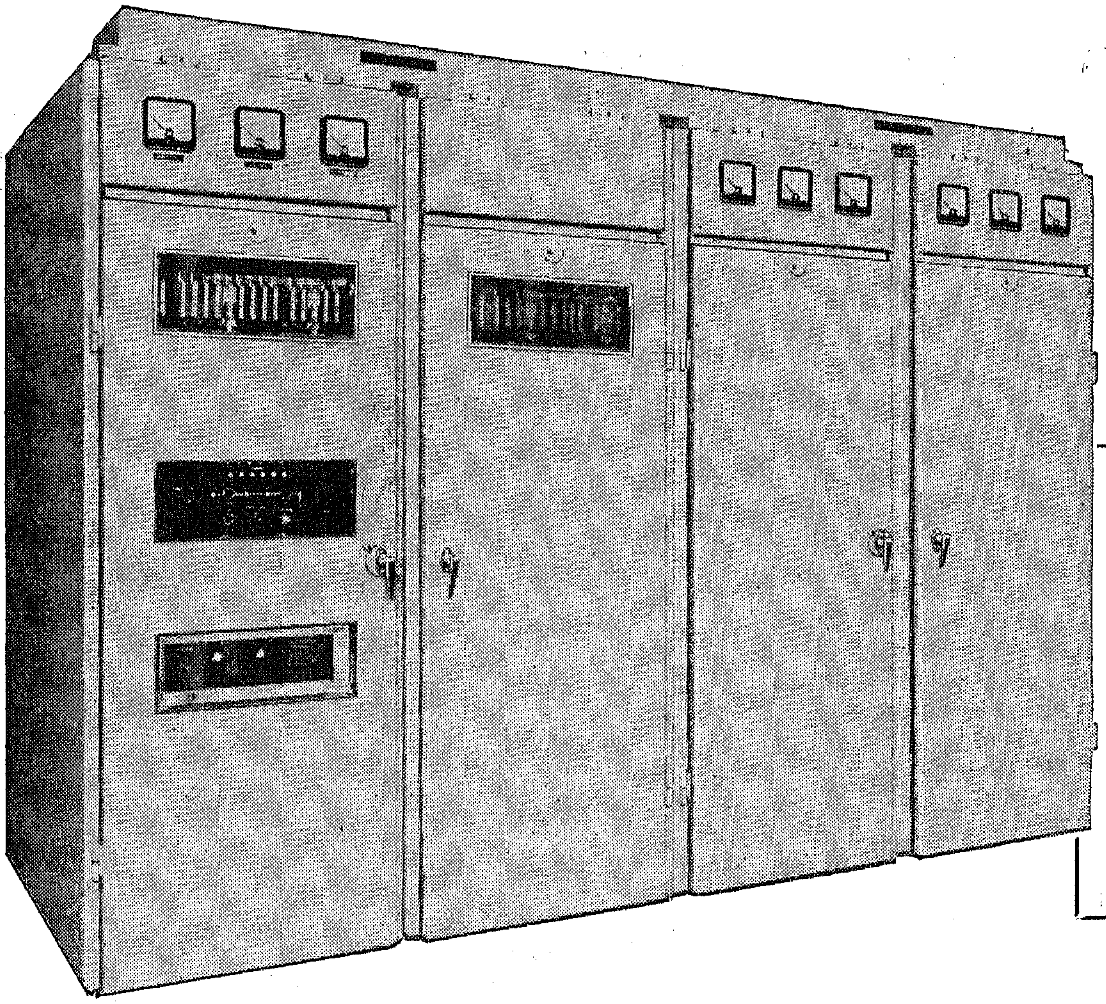


We shall be pleased to discuss this equipment with you.

EKCO electronics

EKCO ELECTRONICS LTD · EKCO WORKS · SOUTHEND-ON-SEA · ESSEX

Marconi 6kW HF ISB Transmitters



TYPES HS 71 AND HS 72

The assembly is enclosed by unit sections, as shown here, with access through front and rear doors. The two left hand bays house the rectifier and power equipment and the right hand bays the low power and auxiliary transmitting circuits and the main output stage.

These transmitters, designed in accordance with the most advanced practice, provide :—

(a) Telegraphy on CW and FSK (A1 and F1)
(b) Independent Sideband Operation (A3b)

The drive equipment is external and provides either ISB modulation or telegraph keying at 3.1 Mc/s and suitable RF oscillator signals for frequency changing in the transmitter. HS 71 is manually operated; HS 72 provides full automatic tuning and selection of any one of six pre-set frequencies.

FEATURES INCLUDE

- **Tuning over the whole range without change of components**
- **Air cooling throughout, with dust filtering.**
- **Double screening of power stages reduces indirect radiation and cooling air noise.**
- **Envelope feed back to reduce distortion.**
- **Compact assembly with good access for servicing and safety interlocking.**

More than 80 countries now have Marconi equipped telegraph and communication services, many of which, completed 20 years ago, still give trouble-free operation.

Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

Surveyed, planned, installed, maintained

Partners in progress with The 'ENGLISH ELECTRIC' *Company Limited*

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

LC12

Reception on six spot frequencies in the HF band and continuous tuning throughout the entire range, plus broadcast reception

MARCONI

RECEIVER

TYPE CR 150/6

The performance of this receiver is of the highest order and meets the requirements of commercial telecommunication working in all climates and conditions. It is of double superhet. design and incorporates special filters, a noise limiter and a built-in, crystal controlled calibration oscillator. H.T. voltages are stabilised to overcome mains fluctuations.

SPECIAL FEATURES

- **Crystal control on any six spot frequencies throughout the band with continuous tunable L.C. oscillator in addition.**
- **Double crystal band-pass filters giving extremely good adjacent channel protection.**
- **Built in 500 kc/s crystal oscillator facilitates calibration checking.**
- **De-sensitising circuit enables full or partial muting when working with an associated transmitter.**
- **Power supply circuits in separate unit to avoid temperature changes.**
- **Suitable for cabinet or rack mounting, with easy servicing access.**

Over 80 countries now have Marconi equipped telegraph and communications systems. Many of these are still giving trouble free service after more than than twenty years in operation.

Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

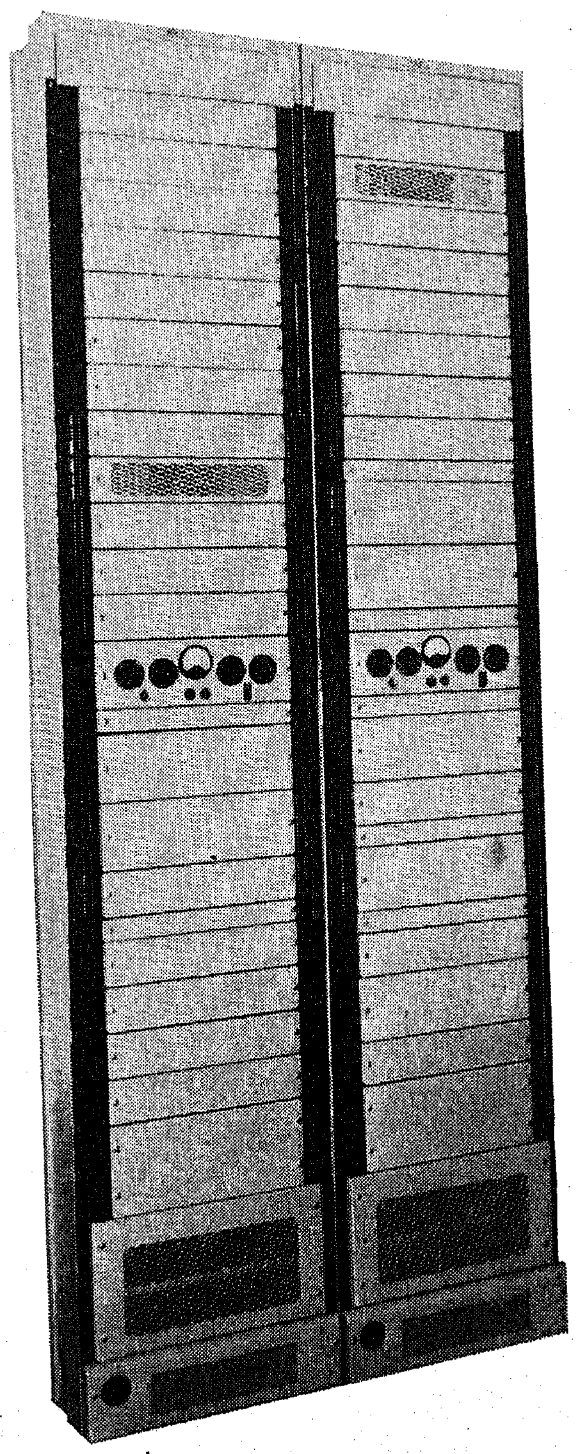
Surveyed, planned, installed, maintained

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

Partners in Progress with The 'ENGLISH ELECTRIC' *Company Ltd.*

LC 9

G.E.C. UHF TELEVISION



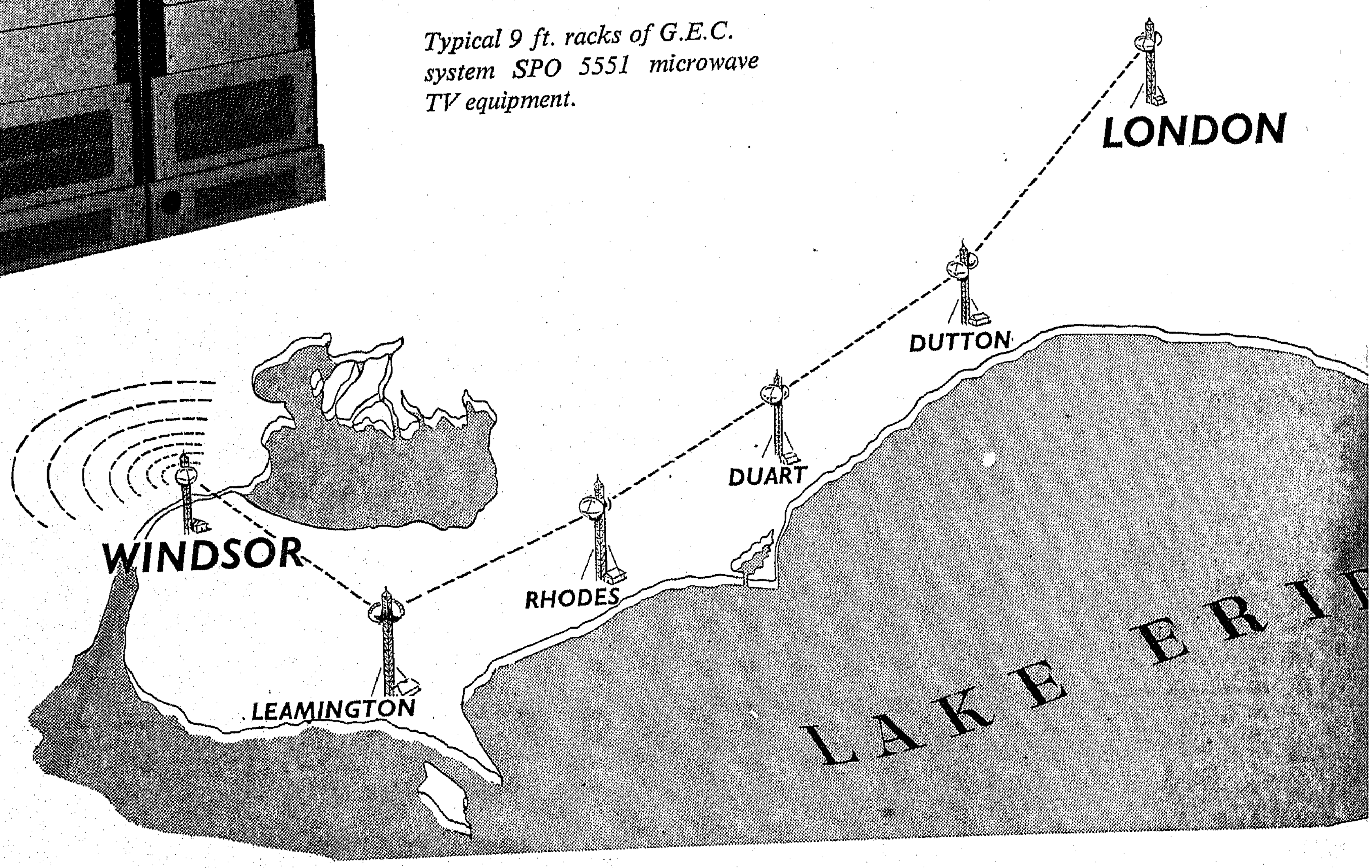
G.E.C. microwave TV equipment has been chosen by Canadian National and Canadian Pacific Railroads for their London-Windsor (Ontario) link.

Canadian Broadcasting Company TV programmes are picked up at London and relayed to Windsor, 120 miles away, through repeater stations at Dutton, Duart, Rhodes and Leamington.

This one-way channel is already in service. The complete installation, proceeding to a firm delivery schedule, looks forward to two further stages. A second, reversible channel is being installed first, and then the original channel is to be made reversible. Supervisory equipment is being supplied for unattended operation of the complete two reversible links.

The TV equipment is of the standard G.E.C. pattern, system SPO 5551 operating in the 1700–2300 Mc/s band. It is suitable for the transmission of 405-line, 525-line or 625-line pictures.

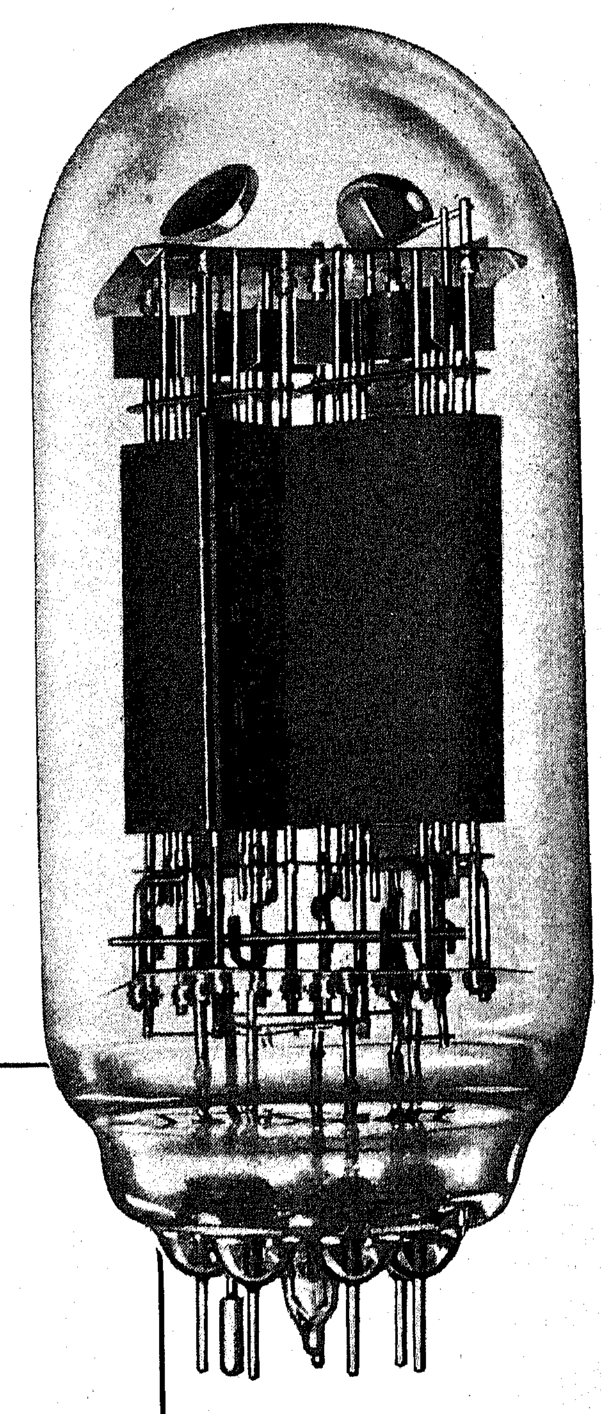
Typical 9 ft. racks of G.E.C. system SPO 5551 microwave TV equipment.



New beam tetrode D.C. control valve of exceptional performance

gm = 35mA/V

Max. cathode current 800mA



The Ediswan Mazda 13.E.1 is a new beam tetrode with a high slope and good power handling capacity for use as either a series or shunt control valve in stabilised power supplies. It is also eminently suitable for servo control motor systems.

In either of these functions the 13.E.1 can usually be used in place of two or three smaller valves thereby saving space and simplifying wiring because multiplicity of connections, grid and anode stopper resistors etc., are avoided, and this, in turn, gives the additional advantage of improved circuit stability.

The 13.E.1 has a B.7A. all glass base and is intended for vertical mounting. All maximum ratings shown below are absolute values, not design centres.

RATING

Vh	26.0 v	13 v.
Ih	1.3 A	2.6 A
Va max	800 V	
Vg2 max	300 V	
Vgl max	−100 V	
Wa max	90 W	
Wg2 max	10 W	
Ik Max	800 mA	

Vh/k max. (cathode+VE) 300 V

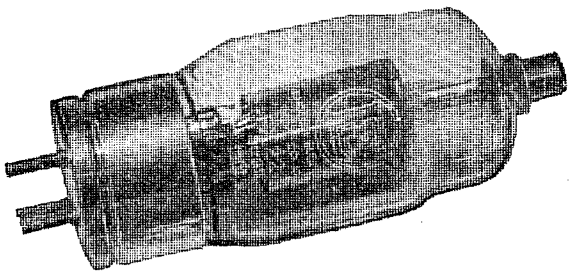
B.7A. BASE CONNECTIONS

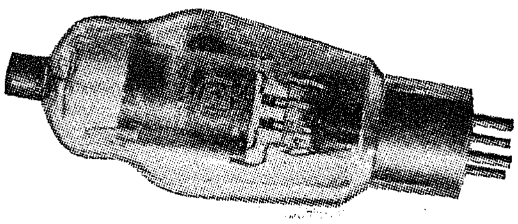
Pin 1	h
Pin 2	h tap
Pin 3	gl
Pin 4	k
Pin 5	g2
Pin 6	a
Pin 7	h

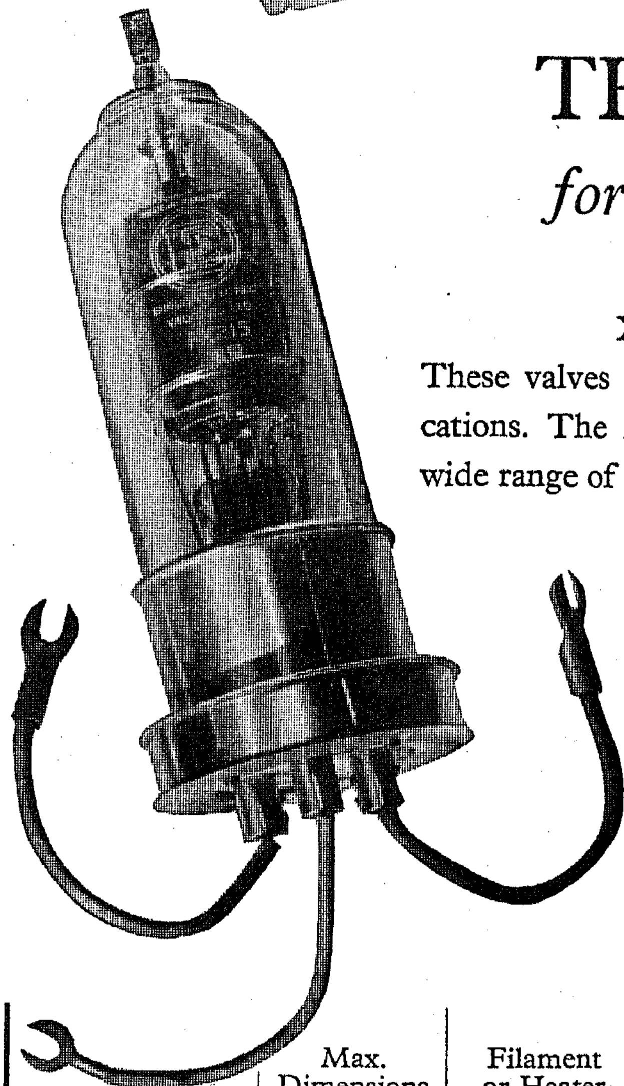
EDISWAN
MAZDA

VALVES AND CATHODE RAY TUBES

THE EDISON SWAN ELECTRIC CO. LTD. 155 Charing Cross Road, London, W.C.2.
Phone: Gerrard 8660. · *Member of the A.E.I. Group of Companies* · Telegrams: Ediswan, Westcent, London







THYRATRONS

for Industrial Purposes

XENON-FILLED TRIODES

These valves are suitable for a wide variety of control applications. The inert gas filling ensures stable operation over a wide range of ambient temperatures.

HYDROGEN-FILLED PULSE MODULATOR TRIODES

Designed to discharge pulse-forming networks in high power, high voltage pulse generators. Short deionization time and low time jitter provide for precise triggering at high repetition frequencies.

Technical data sheets are available on request.

Type	Filling	Max. Dimensions		Filament or Heater		Max. Peak-Inverse Voltage (Volts)	Max. Peak-Forward Voltage (Volts)	Max. Peak Current (Amps)	Max. Mean Current (Amps)	Approx. Tube Drop (Volts)	Peak Power Level (a)	American Equiv.	C.V. Number
		Length (MM)	Dia. (MM)	Volts	Amps (Approx.)								
AFX.212	Xe	54	19	6.3	0.25	350	350	0.11	0.025	11	—	6D4	1949
AFX.203	Xe	176	57	2.5	4.0	300	280	1.7	0.40	11	—	C1A	2868
FX.215	H_2	286	97	2.5	27.5	16,000	16,000	200	0.20	100(b)	2.0×10^9	—	2203
FX.219	H_2	222	65	6.3	10.6	16,000	16,000	325	0.20	100(b)	3.2×10^9	5C22	2520
FX.225	H_2	175	65	6.3	6.1	8,000	8,000	90	0.10	100(b)	2.0×10^9	4C35	1787
FX.227	H_2	132	40	6.3	2.25	3,000	3,000	35	0.045	100(b)	0.3×10^9	3C45	372

Xe—Xenon H_2—Hydrogen

(a)—Product of Peak forward Voltage, Peak Current and pulse repetition frequency.

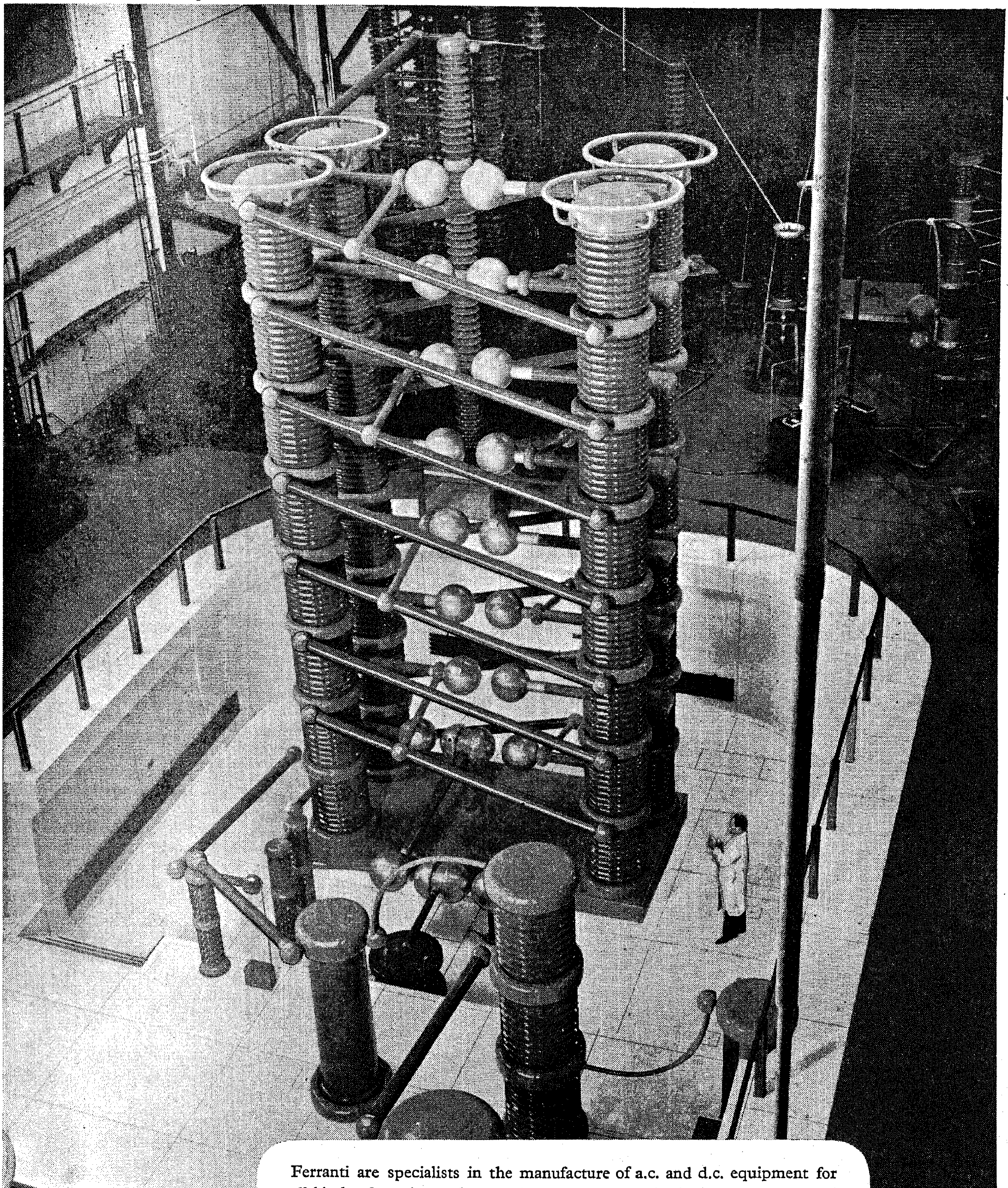
(b)—Under conditions of pulse operation.

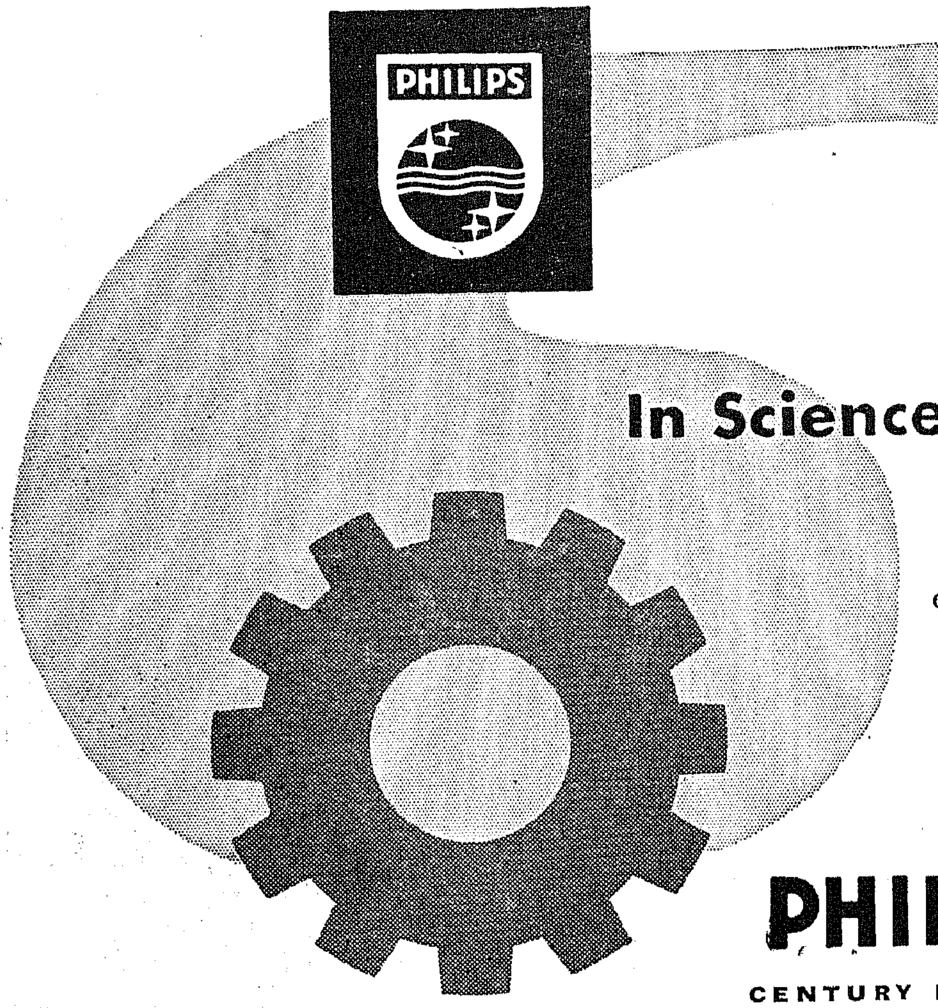
'ENGLISH ELECTRIC'

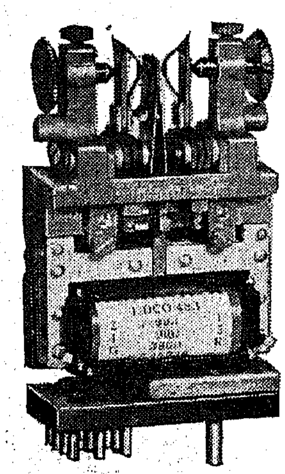
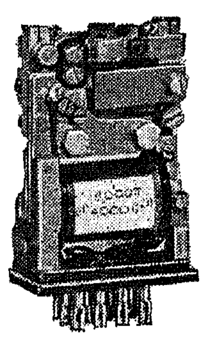
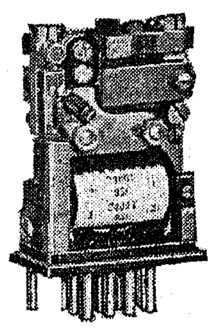
ENGLISH ELECTRIC VALVE CO. LTD.

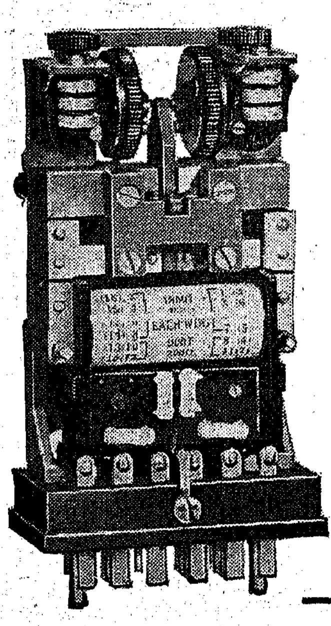
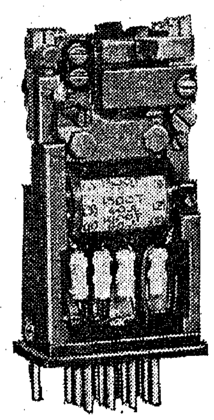
Waterhouse Lane, Chelmsford
Telephone: Chelmsford 3491

AP300/13




PHILIPS

Lines of communication . . .

In June 1920 the first advertised sound radio programme was broadcast from the Marconi transmitter at Chelmsford. To-day Marconi high or medium power transmitters and high power aerials are installed in every one of the B.B.C's television transmitter stations and Marconi television systems are being supplied to countries in North and South America, Europe and Asia.

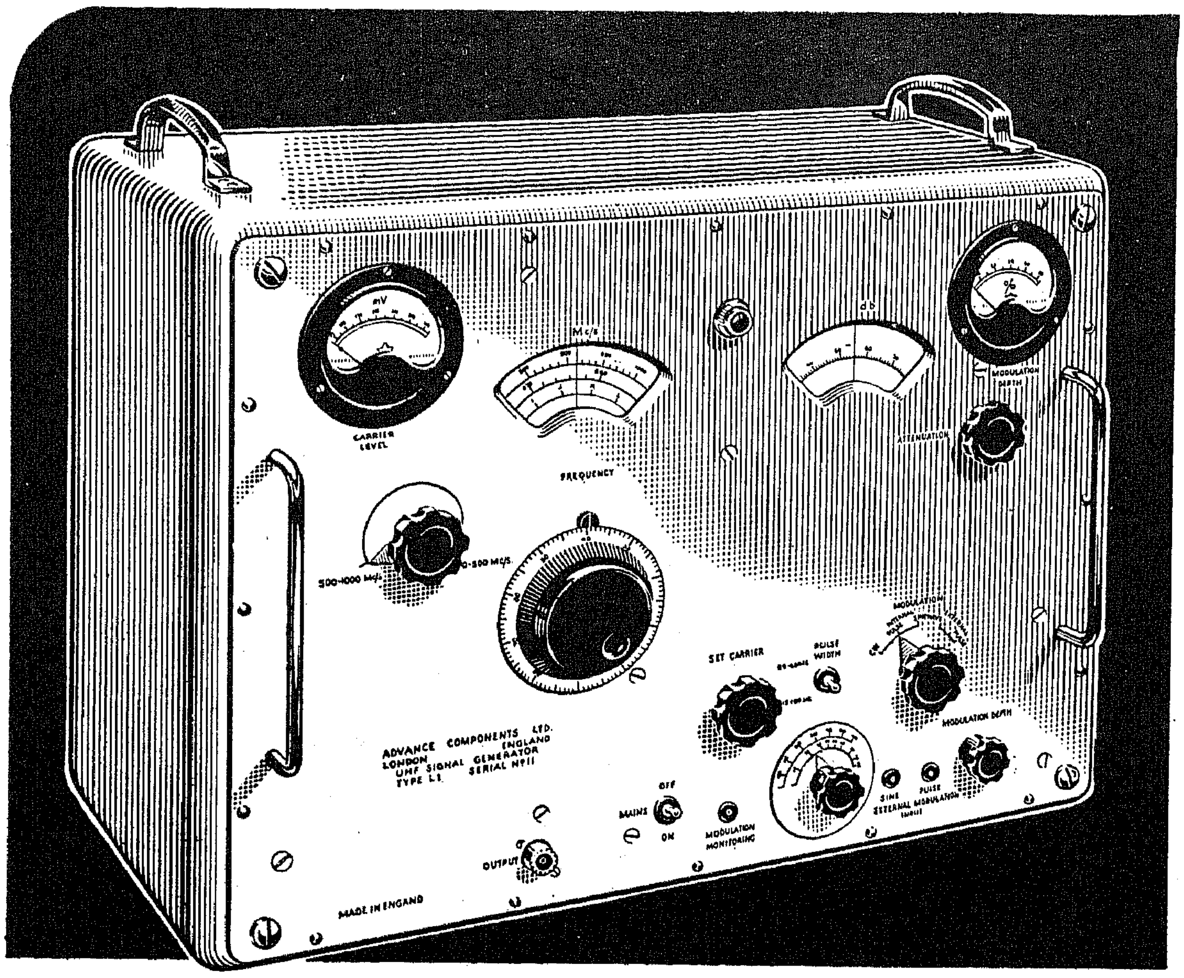
TELEVISION CAMERAS
STUDIO EQUIPMENT
RADIO LINKS AND TRANSMITTERS
AERIAL SYSTEMS
COLOUR TELEVISION
INDUSTRIAL TELEVISION

MARCONI

Complete Television and Sound Broadcasting Systems

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · CHELMSFORD · ESSEX

LG 3



MODEL L1 — 300 TO 1,000 Mc/s

Full technical details in Folder A25 available on request.

Operating in the decimetre wave region, this Signal Generator has been designed to meet the increasing requirements of the development laboratory at Ultra High Frequencies. Besides a wide frequency range and reliable attenuator system, provision is made in the L.1 for both internal and external sinewave and pulse modulation. Full technical details are available on request.

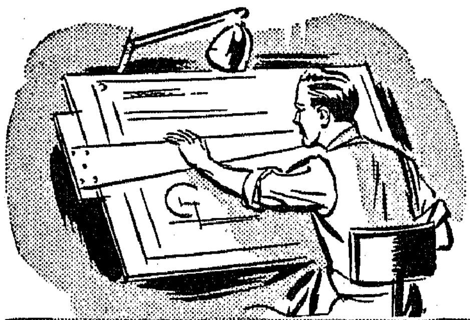
Advance

U.H.F. SIGNAL GENERATOR

Nett price in U.K. £375

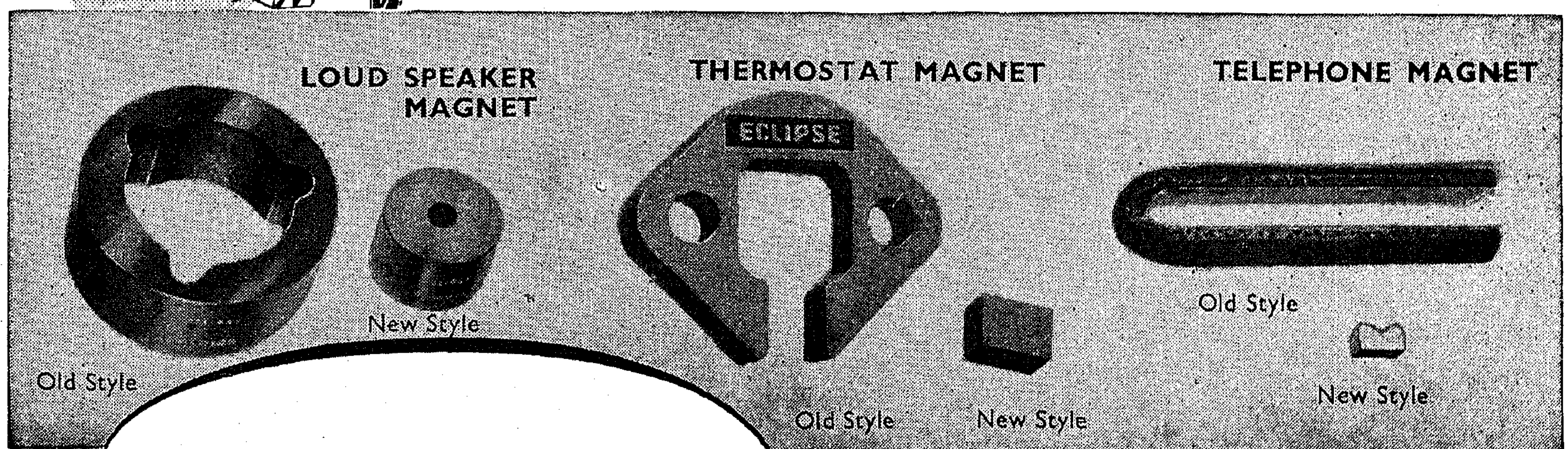
ADVANCE COMPONENTS LTD. MARLOWE RD., WALTHAMSTOW, LONDON, E.17. Phone: LARkswood 4366/7/8

G.D.



Economy begins at the design stage . . .

The latest developments in permanent magnet materials and techniques can effect remarkable economies, provided they are introduced at the design stage.



Send for Publication P.M.131/51 "Design and Application of Modern Permanent Magnets"

Made by the makers of "Eclipse" Permanent Magnet Chucks

Eclipse PERMANENT MAGNETS

JAMES NEILL & CO. (SHEFFIELD) LTD.
SHEFFIELD 11 ENGLAND

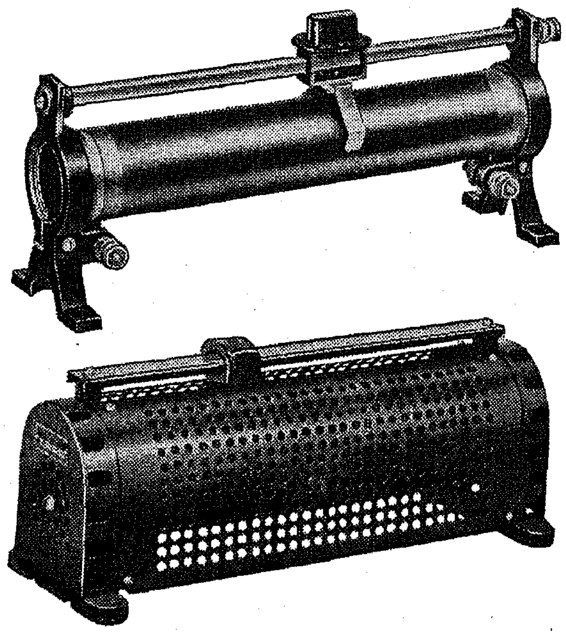
M.L.

ZENITH

(REGD. TRADE-MARK)

TUBULAR SLIDING RESISTANCES

Zenith Resistances of proved durability are in constant satisfactory use in all parts of the world. They are available in a great variety of types and sizes, and are ideal for use in laboratories and test rooms.



Illustrated catalogue of all types free on request

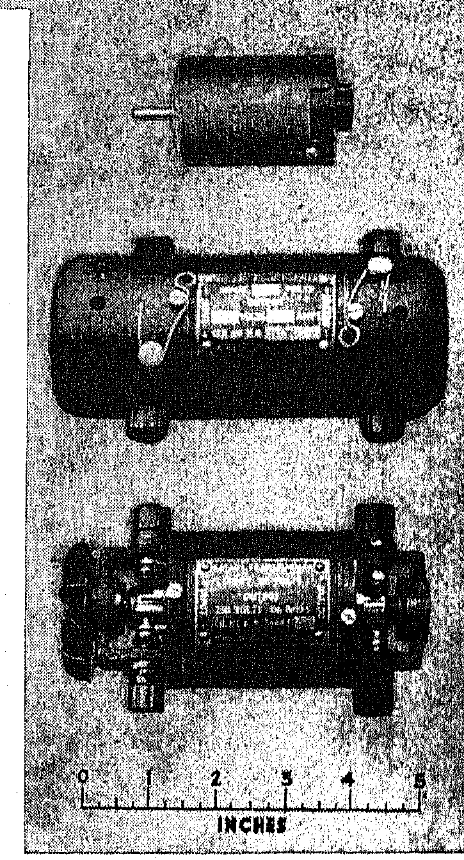
The ZENITH ELECTRIC CO. Ltd.
ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREEN
LONDON, N.W.2

Telephone: WILlesden 6581-5 *Telegrams:* Voltaohm, Norphone, London

MANUFACTURERS OF ELECTRICAL ENGINEERING PRODUCTS INCLUDING RADIO AND TELEVISION COMPONENTS

NEWTON-DERBY

LIGHTWEIGHT ELECTRICAL EQUIPMENT



Our manufactures include:

- Aircraft Generators and Motors
- Automatic Voltage Regulators
- Rotary Transformers
- High Frequency Alternators
- H.T. D.C. Generators

The illustration shows a small high-speed fractional H.P. motor and miniature Rotary Transformers for "Walkie Talkie" and other RADIO applications.

NEWTON BROTHERS (DERBY) LTD.

HEAD OFFICE & WORKS: ALFRETON ROAD, DERBY.
TELEPHONE: DERBY 47676 (3 LINES) TELEGRAMS: DYNAMO, DERBY.
LONDON OFFICE: IMPERIAL BUILDINGS, 56 KINGSWAY W.C.2.

THE PROFESSIONAL ENGINEERS APPOINTMENTS BUREAU

Incorporated under Limited Guarantee, 1947
Licensed annually by the London County Council

9, Victoria Street, London, S.W.1

The Bureau invites applications for registration for employment from members who, by reason of their engineering qualifications, belong to The Institution of Civil Engineers, The Institution of Mechanical Engineers, or The Institution of Electrical Engineers or persons whose engineering qualifications for election or admission to one of those Bodies have been approved by the respective Councils. The necessary forms may be obtained on application to the Registrar of the Bureau; a stamped addressed foolscap envelope should be enclosed.

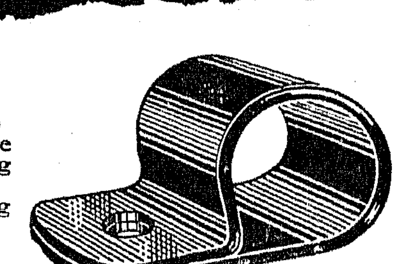
Employers of Professional Engineers are invited to notify vacancies, and to assist the Bureau by furnishing the following particulars:

1. Title of post with description of duties.
2. Age range. *This should be made as wide as possible, and it should be noted that there are many experienced applicants in the upper age groups.*
3. Salary, and any supplementary remuneration.
4. Professional, technical and special qualifications such as foreign languages.
5. Location of work. *Notifications of vacancies in the Dominions are especially welcome.*
6. Could a successful candidate be assisted in finding suitable living accommodation locally?
7. Does the post involve the control of staff? If so, how many technicians and/or other workers?
8. Any other information which will assist in the selection of candidates.

PLASKLIP *secures cables*

Quicker, Simpler, Safer

The Plasklip has the advantage of being non-metallic and impervious to tropical conditions, it is the answer to all cable securing problems in electrical equipment. Made of strong non-magnetic high dielectric material it will outlast the cables or components it secures.

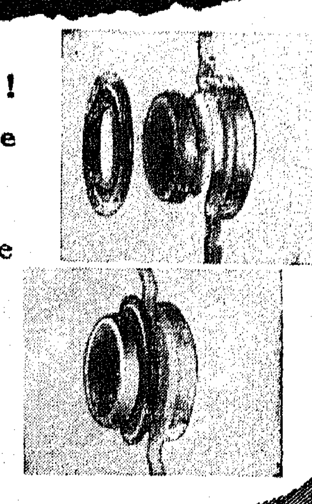


the snap on Bush

that can't come out!

Easy to fix Safe in use
Saves time

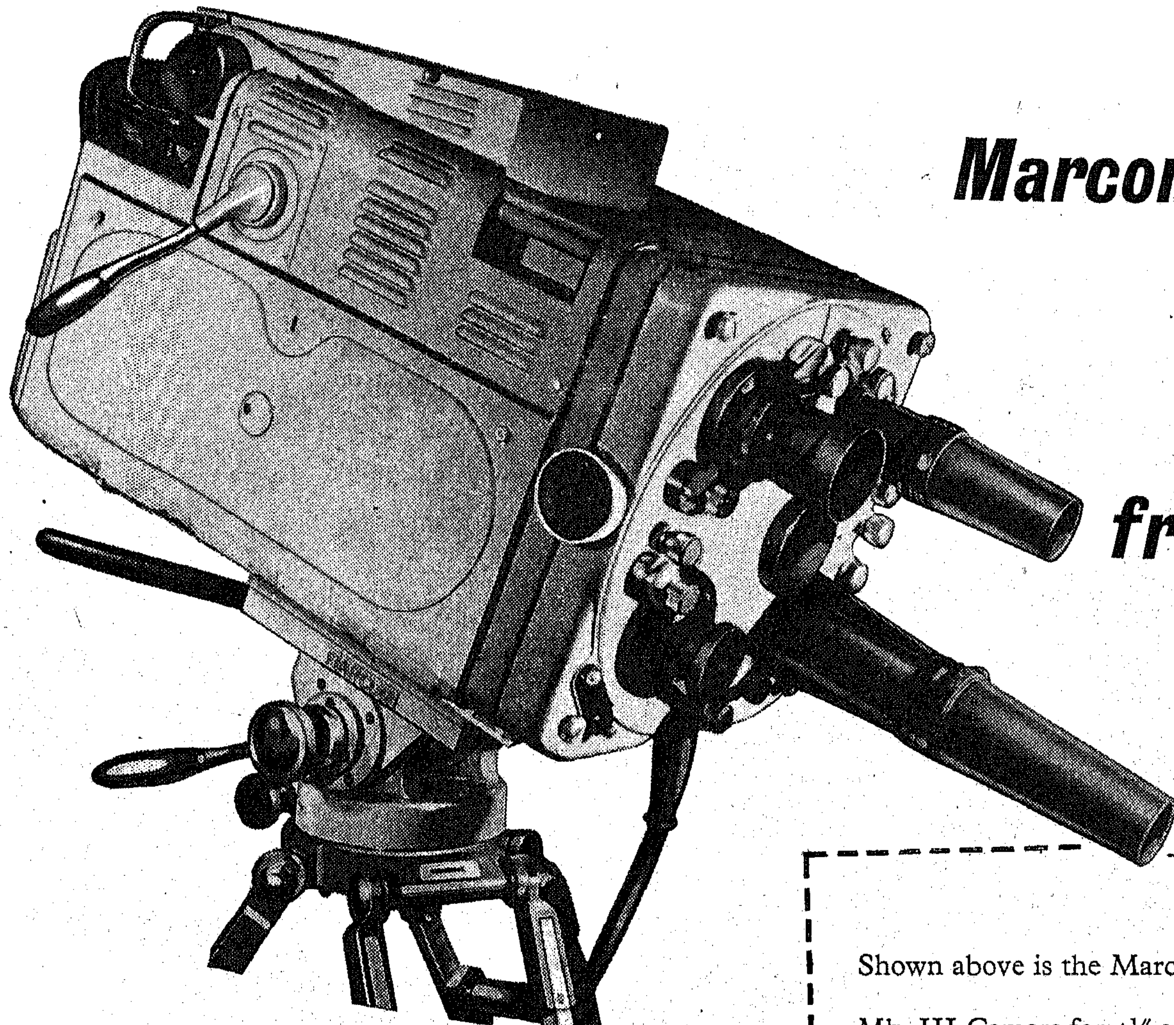
The Insuloid Bush is instantly fixed by a simple finger action, and yet it provides complete security when assembled. High dielectric. Fully tropical.



The PLASKLIP and BUSH are approved by all services

Samples and literature available on request

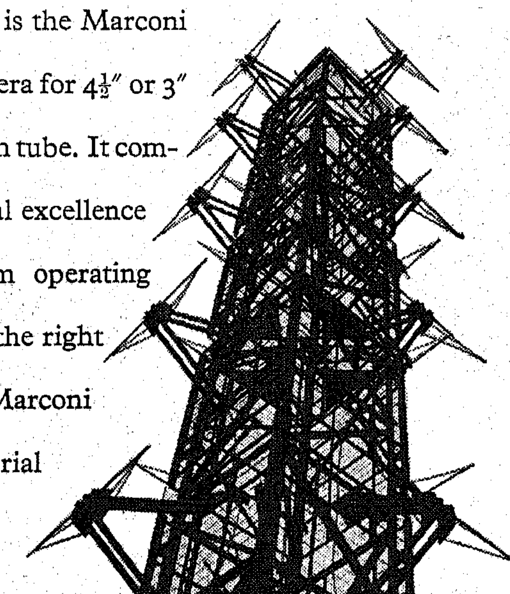
INSULOID MANUFACTURING COMPANY LIMITED
Sharston Works, Leaston Ave., Wythenshawe, Manchester.
Tel.: Wythenshawe 2842



Marconi Complete TELEVISION SYSTEMS from Camera to Aerial

Both for studio and outside broadcasting of television programmes Marconi equipment is required at every stage of production and transmission. Cameras, Picture and Waveform Monitors, Vision Mixers, Telecine Equipment where film sequences are included, Microwave Links, Transmitters and Aerial must all be matched as components of a completely integrated system with a designed performance. The long experience and advanced technical knowledge of the Company's Broadcasting Division are at the disposal of all who are responsible for television, throughout the world.

Shown above is the Marconi Mk. III Camera for $4\frac{1}{2}''$ or $3''$ image orthicon tube. It combines technical excellence with optimum operating facilities. On the right is a typical Marconi Television Aerial array.



Marconi Television Equipment is installed in every one of the B.B.C.'s Television stations and has been supplied to countries in North and South America, Europe and Asia. Compatible colour television was first demonstrated in Britain by Marconi's.

Lifeline of communication

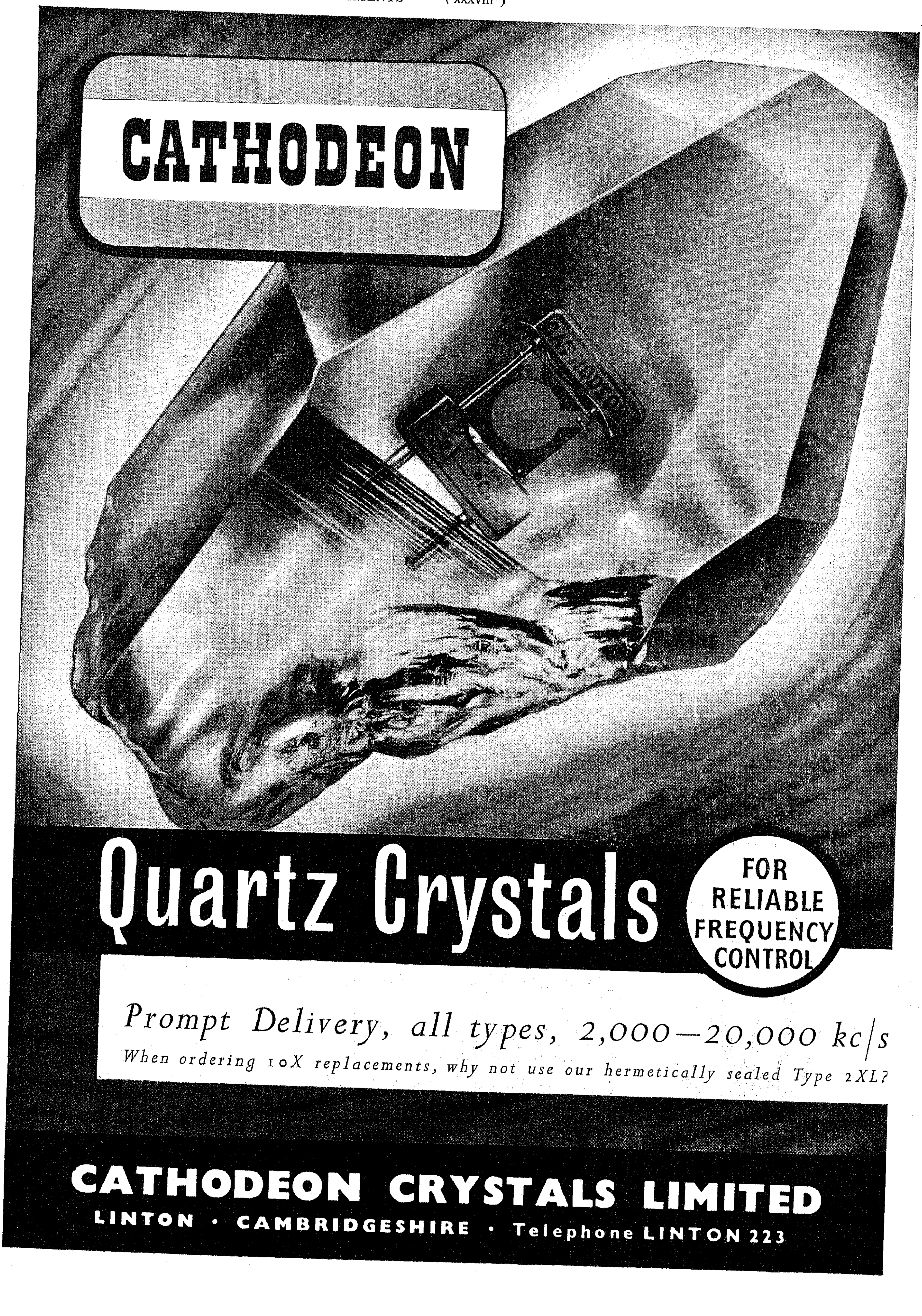
MARCONI

Complete Broadcasting and Television Systems

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

Partners in progress with the 'ENGLISH ELECTRIC' *Company Limited*

LB 5


CATHODEON
Quartz Crystals
FOR RELIABLE FREQUENCY CONTROL
Prompt Delivery, all types, 2,000—20,000 kc/s
When ordering 10X replacements, why not use our hermetically sealed Type 2XL?
CATHODEON CRYSTALS LIMITED
LINTON · CAMBRIDGESHIRE · Telephone LINTON 223

G.E.C.

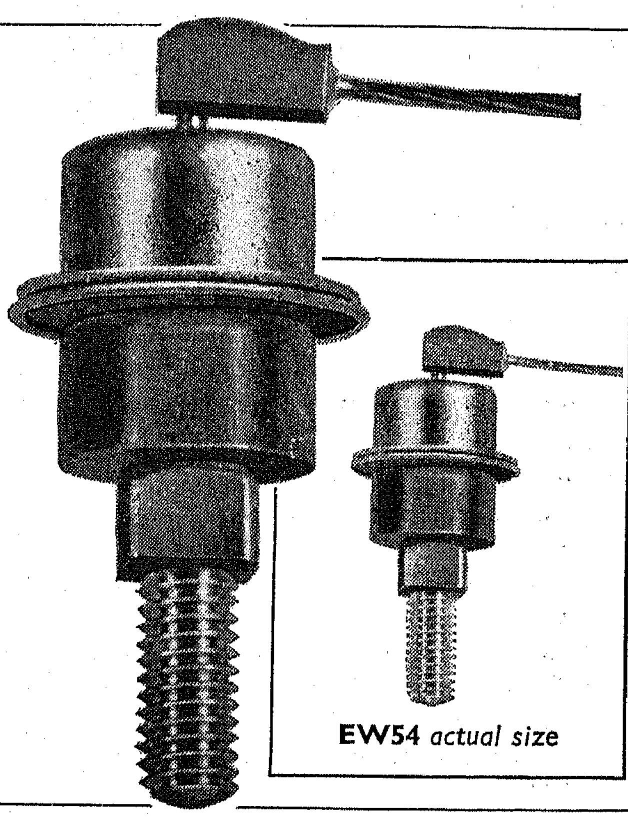
EW54 Germanium Junction Power Diode

NOW AVAILABLE TO ELECTRONIC EQUIPMENT MANUFACTURERS

The EW54 p-n junction germanium diode is intended for use in rectifier circuits at medium voltage and current. The diode is hermetically sealed in a copper container. This is particularly important because of the deleterious effects of moisture on germanium devices.

The main features of this type of diode are high rectification efficiency and small size; the former results primarily from the very low forward resistance of the diode.

The diode is of value in applications requiring outputs up to the order of 20A at voltages up to 50 (at 20°C), using a full-wave bridge arrangement.



EW54 *actual size*

For a typical diode, at an ambient temperature of 20°C:

Current at +0.5V=8A
at —100V=6mA

Other recent G.E.C. Semi-Conductors are the

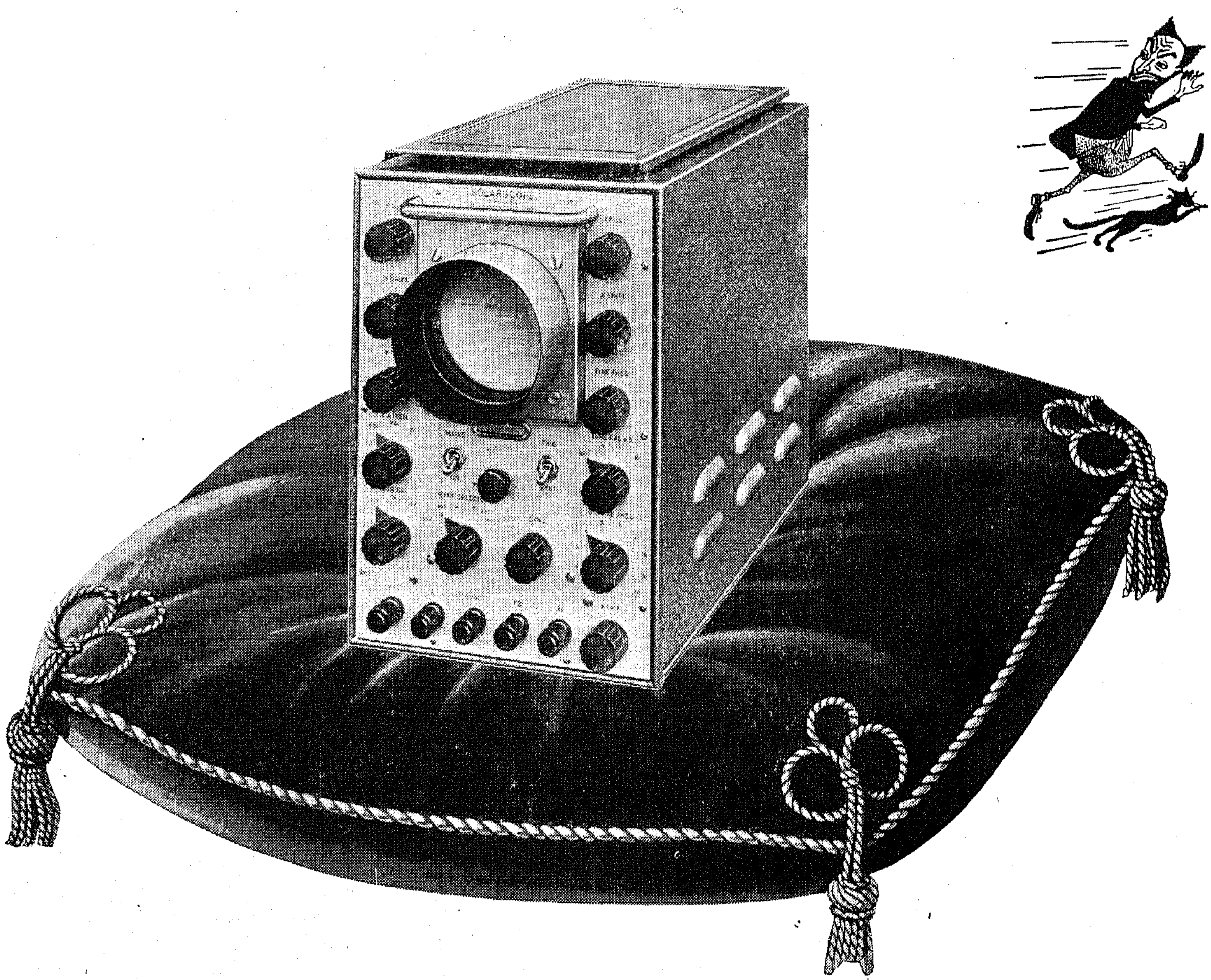
EW53 EW58 EW59

Details of all the above devices may be obtained from

THE OSRAM VALVE AND ELECTRONICS DEPARTMENT

THE GENERAL ELECTRIC CO. LTD., MAGNET HOUSE, KINGSWAY, LONDON, W.C.2

APPLICATIONS *Unlimited*



The Constant-Bandwidth Solarscope CD.514 has surprised everybody except us. It has attracted outstanding attention at the Physical Society and Instrument Industries Exhibitions, and the volume of orders already received indicates its immense field of applications. The CD.514's high sensitivity and ability to display full screen pictures at the higher frequencies, coupled with its remarkably low price of £105, has established for itself an undisputed lead over all competitors. If you've missed out on it so far, get in touch with us now.

BRIEF SPECIFICATION OF THE CONSTANT-BANDWIDTH CD. 514

"Y" Sensitivity: 30 mV/cm to 30 V/cm from 5 c/s to 9 Mc/s (3 db)
"Y" Calibration: 50 cycle sine-wave, 100 mV, 1 V, 10 V and 100 V
Input Impedance: 1 Megohm in parallel with less than 30 pF
"X" Amplifier
Bandwidth: 1.5 c/s to 900 Kc/s (3 db)
"X" Expansion: 0.5 to 5 diameters
Time Calibration: 0.1 microsec., 1 microsec. & 10 microsec. markers
T.B. Frequency: 15 c/s to 300 Kc/s.
T.B. Trigger
Sensitivity: Iv. R.M.S., minimum
Other Features:
Post-deflection-acceleration tube giving brighter trace.
Direct access to tube deflection plates.
Illuminated graticule, adjustable stand.
Takes standard camera attachments.
Weight: 40 lbs. Size: 12½" × 8¾" × 16½" deep overall

THE SOLARTRON ELECTRONIC GROUP LTD.

RELIABILITY—UNDER OUR 12 MONTHS' GUARANTEE COSTS HAVE NEVER EXCEEDED 0.2% OF SALES

Thames Ditton, Surrey
Telephone: EMBerbrook 5611
Cables: Solartron, Thames Ditton

THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 102. PART B. No. 4. JULY 1955

621.385.832

Paper No. 1883 R
July 1955

SOME HALF-TONE CHARGE STORAGE TUBES

By R. S. WEBLEY, B.Sc., H. G. LUBSZYNSKI, Dr.Ing., F.Inst.P., and J. A. LODGE, B.Sc., Graduate.

(*The paper was first received* 11*th November*, 1954, *and in revised form* 23*rd April*, 1955.)

SUMMARY

Four tubes are described which have been developed for use primarily in bright radar display systems. Two of the tubes are full-storage television camera tubes in which sufficient charge is stored to enable many reading scans to be performed before the charge is consumed. In the other two tubes a charge pattern is used to modulate an electron beam, the charge pattern itself remaining unchanged in the process. In the two former tubes and in one of the latter, the information to be stored must be supplied in the form of an optical image. In the fourth tube it must be supplied as an electrical signal.

Continuous reading times ranging from a few seconds to several hours have been achieved. Resolutions of up to 700 spots per picture diameter are also reported. With some of the tubes, the erasure rates are readily controllable over a wide range.

(1) INTRODUCTION

Storage tubes are comparatively recent newcomers to the range of electronic devices, but they are rapidly finding application in many fields, including radar, computation and medicine and also as research tools. These tubes owe much of the knowledge used in their design to the research on television camera tubes, which in the past quarter of a century has contributed a great deal to high-vacuum electronics.[1,2,3]

The tubes to be described have been given the C.V.D. designations VCRX343, 360, 326 and 350 (see Fig. 13). They were developed for use in radar display, and were designed to store half-tone pictures, giving long storage times with continuous reading.

(2) GENERAL

In all the tubes to be described the information is stored in the form of a charge pattern on an insulating or semi-conducting surface. The information is written in, read out, and erased by means of electron beams, photo-emission or photo-conduction. The electron beams may be generated either by an electron gun and scanned by conventional means, or by scanning a light spot in the appropriate raster over an extensive photo-cathode.

The type of tube and its precise specification depend greatly on the application for which it is to be used. However, certain factors are common to all types of tube. Spurious signals, such as shading and background imperfections, must always be sufficiently small not to be confused with stored information, or to degrade the quality of the stored information appreciably. The output signal must also be large enough to be distinguishable from random noise. The resolution of the tube must be sufficient to store all the information supplied to it and to give out this information at a later time, with an acceptable amount of degradation. The tube should also be able to record information at the highest rate at which it may be supplied.

Generally, in picture storage tubes, it is desirable to be able to control over a wide range the time for which a picture can be read, the limits being fixed by the particular application. This time may vary from $\frac{1}{30}$ sec to 3 hours. The transfer characteristic should usually be as linear as possible, and often a linear integration range is also required. The geometry of the reproduced picture should be close to that of the original. Other properties, such as absence of coupling between input and output, and ability to write and read simultaneously with asynchronous scanning systems, may also be of importance.

(3) TUBES OPERATING WITH CHARGE RESTORATION

(3.1) Thin-Target C.P.S. Emitron Type VCRX343

(3.1.1) General.

A multiplier orthicon using a special thin target was first described by Forgue[1] for long-term storage. A C.P.S. Emitron[2] with a similar target was later developed by the authors for use in a bright p.p.i. display system. A block schematic of the system is shown in Fig. 1. The lag normally associated with the afterglow phosphor was incorporated in the scan convertor tube.[1,4]

The tube is now superseded in this application, but it is described briefly because information discovered during the development influenced subsequent work.

(3.1.2) Description.

The basic C.P.S. Emitron has been described fully elsewhere.[2] For this application, the capacity of the target was increased by a factor of 16 over that of the standard C.P.S. Emitron. This was done by using a disc of mica 0·0002–0·0003 in thick, and probably this still represents the practical limit with photo-emissive tubes. On this target a charge image of the information to be stored was formed by means of a cathode-ray-tube lens system, the phosphor and photo-emitter being matched for colour characteristics.

The authors are at the E.M.I. Research Laboratories, Ltd.

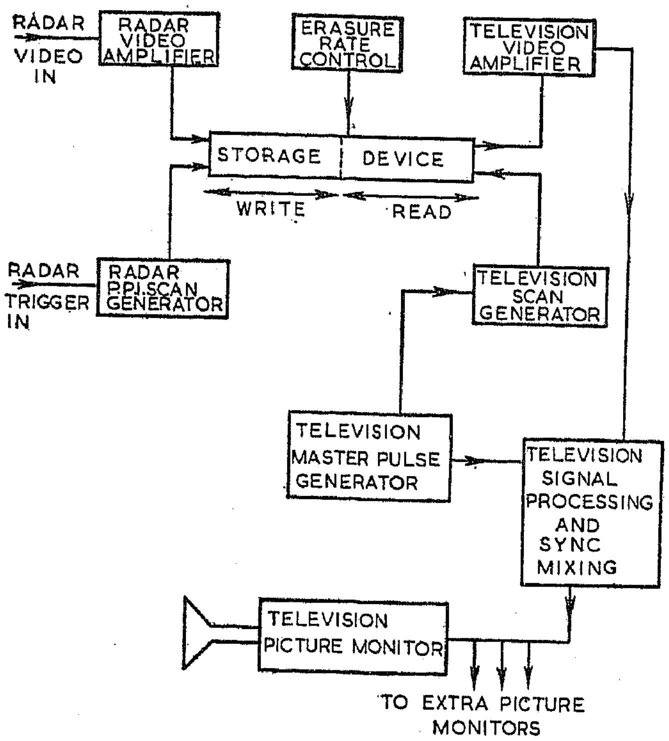


Fig. 1.—Block schematic of storage-tube radar display system.

The basic system of Fig. 1 was utilized to test the tubes, using the standard British television frequencies in the reading channel. The reading time was found to be 8 sec when starting with an initial signal/noise ratio of 12 dB, and with this time the limiting resolution was about 400 lines (television standard).*

These values were obtained due to a variety of reasons. In the first place, the amplifier was of comparatively early design. Using a modern type of amplifier,[5] a reading time of about 15 sec with an initial signal/noise ratio of 24 dB and some improvement in resolution are to be expected. In addition, advances are currently being made in the focusing and scanning arrangements of the standard C.P.S. Emitron, and these are applicable to types VCRX343, 350, and 360.

In order to obtain the maximum possible reading time, the beam current of type VCRX343 was reduced to the limit set by an acceptable output signal/noise ratio. Also a larger voltage swing was permitted on the mosaic, and the cathode-ray tube was driven harder than is the usual practice. Each of these factors contributed to a loss of resolution. It was clear that, with light-operated charge-restoration tubes, improvements could only be obtained by developing some new type with a considerably increased capacity in order to reduce the voltage swing and hence the defocusing of the reading beam. Also a higher light sensitivity was required to avoid over-driving the cathode-ray tube. It was desirable that the tube should be stable with respect to repetitive excitation arising from permanent echoes.

Furthermore, the use of a photo-emitter is undesirable on two accounts. The effective light sensitivity was found to be lower for the small illuminated areas which were of primary interest than for the larger areas. This effect, known as coplanar biasing, is illustrated in Fig. 2, in which the potentials of two illuminated areas of differing size were measured as a function of exposure, using the electron beam as a probe. Also, spurious signals resulted from the emission of photo-electrons. This effect was first reported by Forgue.[1] It was found possible to reduce the spurious signals to negligible proportions by the use of a cathode-ray-tube phosphor with an afterglow of some 50 millisec. Complete cancellation was obtained by using a subsidiary photocell on which a proportion of the emitted light was incident.

It will be seen later that the photo-conductive tube type

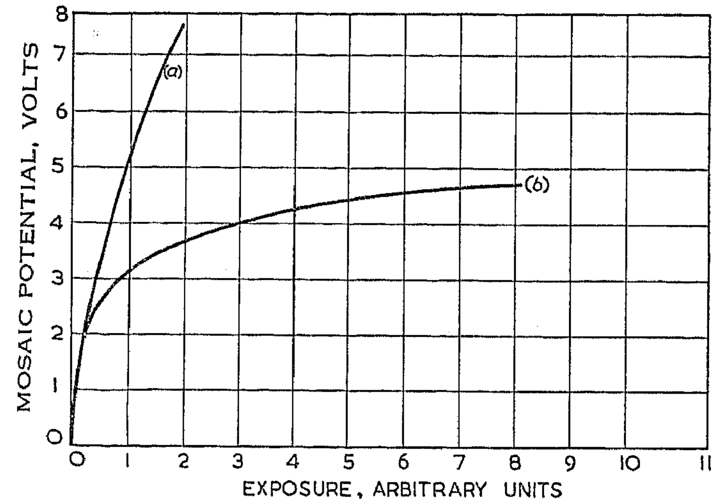


Fig. 2.—Target voltage as function of exposure for areas of different sizes in VCRX343.

(*a*) Circular area one-twelfth of mosaic diameter.
(*b*) Circular area one-hundredth of mosaic diameter.
Mosaic coverage: 60%.

VCRX360 had all the required properties. With the beam-modulation tube type VCRX350, in which electron beams are used for writing, means were devised to minimize the spurious signals, and coplanar biasing was used to advantage to produce stability.

(3.2) Photo-Conductive Storage Tube Type VCRX360

(3.2.1) General.

As described in Section 3.1.1, a tube working on the charge-restoration principle should have as high a sensitivity and target capacitance as possible. A number of photo-conductive materials have a high apparent quantum efficiency, approaching unity and sometimes even exceeding it. Moreover, efficient photo-conductive target layers can, in general, be made very thin. This, together with a high dielectric constant, results in a target of high capacitance. Therefore, development was started on the photo-conductive storage tube type VCRX360.

(3.2.2) Description.

The tube is of the low-velocity scanning-beam type[2] and is shown diagrammatically in Fig. 3. The signal plate 6 is a semi-

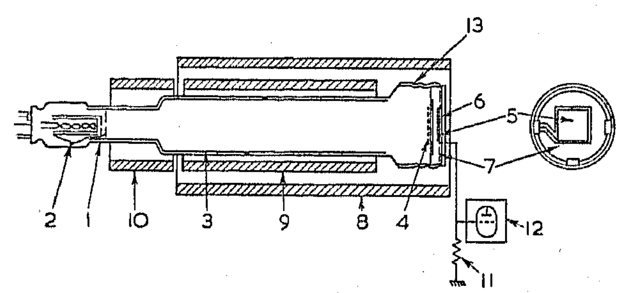


Fig. 3.—Photo-conductive storage tube type VCRX360.

1. Glass envelope.
2. Electron gun.
3. Metallized wall anode.
4. Ion-trap mesh.
5. Target.
6. Signal plate.
7. Signal-plate surround.
8. Focus coil.
9. Scanning coils.
10. Alignment coils.
11. Signal resistor.
12. Head amplifier.
13. Front-wall anode.

transparent conducting coating of tin oxide, applied to the inside surface of the end window. It is surrounded by a metallized layer, 7, held at the potential of the gun cathode.[7] The photo-conductive layer, 5, is then deposited over the whole of the end window, covering both signal plate and surround. Without the metal surround, the unscanned area of the target gradually leaks to the potential of the signal plate, and thus a non-uniform potential distribution in the plane of the target is set up, causing geometric distortion. The scanned area of the target is 1 in × 1 in. The target layer consists of antimony trisulphide

* Where figures for resolution are given in the paper, they refer to "television lines," i.e. the number of equally wide, vertical, black and white lines over the whole width of the target. For example, a resolution of 400 lines refers to 200 white and 200 black lines.

deposited on the signal plate by evaporation from two channels mounted on a metal frame which carries an ion-trap mesh, 4. The frame is connected to the front part of the wall electrode and also to a contact on the window. The rear portion of the wall electrode, 3, is connected to the pinch. The front portion is decoupled to earth by an external capacitor and resistor. It has been found[8] that in this arrangement the ion-trap mesh acts as an electrostatic screen for the signal plate and eliminates any pick-up on the latter of the large pulses produced by the line-return transients in the scanning coils,[9]. This becomes essential where line-by-line clamping of the black level is to be carried out, because in this arrangement the signal during line-return times is used as the reference black level.

In operation, the signal plate is held at a potential of about 4–10 volts positive with respect to the gun cathode, while the target surface is stabilized at cathode potential by the electron beam. Thus a potential difference is established across the target layer.[9] The target thickness is about 0·5 micron, and the potential gradient across the layer is thus approximately 100 kV/cm. Under illumination the target conductivity increases and the potential of the scanned surface rises towards that of the signal plate. In scanning, the target potential is restored towards cathode potential, and thus a signal is developed in the signal plate circuit.

(3.2.3) Mechanism of Signal Generation and Lag.

The target surface P (Fig. 4) has a capacitance C to the signal plate S. In parallel with C is the resistance R of the

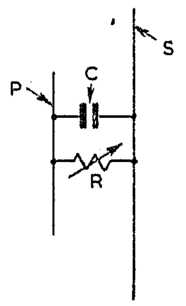


Fig. 4.—Equivalent circuit of target of VCRX360.

target material, which varies as a function of the light falling on it. In the type VCRX360, the time-constant of the layer RC is always large compared with the frame time of $\frac{1}{25}$ sec.*

(3.2.3.1) *Capacitance Lag.*

Of the total beam current I_b approaching the target only a fraction i_b is able to land, the remainder returning towards the gun. The current i_b is a function of the target potential.

Let v be the target potential, and Q its charge.

Then $$i_b = f(v) \qquad (1)$$

and the charge deposited on the target is

$$\delta Q = i_b \delta t = C\delta v$$

Therefore $$\frac{\delta v}{f(v)} = \frac{1}{C}\delta t \qquad (2)$$

If the target is discharged from a starting potential v_0 to a potential v_T during a time T

$$\int_{v_0}^{v_T} \frac{dv}{f(v)} = \int_0^T \frac{1}{C}\, dt = \frac{T}{C} \qquad (3)$$

* In later equipment (see Section 3.2.9) the frame time was increased to $\frac{2}{25}$ sec. This had no apparent effect on the mechanism of decay.

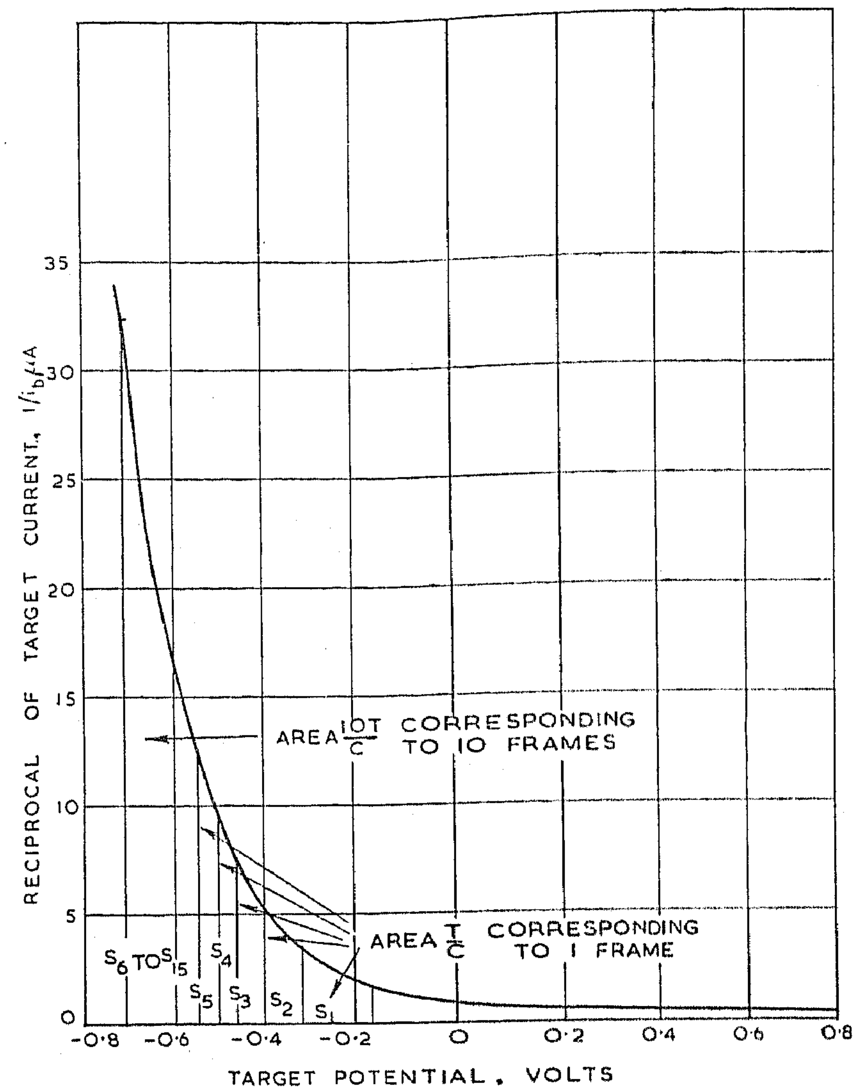


Fig. 5.—Reciprocal of beam-acceptance curve $\frac{1}{i_b} = \frac{1}{f(v)}$. Target constant $T/C = 0{\cdot}4$ volt/μA for $T = \frac{1}{25}$ sec = 1 frame.

First signal	$S_1 = 0{\cdot}4$ μA.
Second signal	$S_2 = 0{\cdot}22$ μA after 1 frame.
Third signal	$S_3 = 0{\cdot}15$ μA after 2 frames.
Fourth signal	$S_4 = 0{\cdot}12$ μA after 3 frames.
Fifteenth signal	$S_{15} = 0{\cdot}031$ μA after 14 frames.

The beam acceptance curve $i_b = f(v)$ was measured in special tubes with conducting targets by Dr. Meltzer and Mr. Holmes. Fig. 5 shows $1/i_b = 1/f(v)$, the reciprocal of the beam acceptance curve. Using eqn. (3), the discharge characteristic of the tube can be constructed. Starting from a potential v_0 at time $T = 0$, the target potential v_T at time T is found by numerical integration, so that the area under the $1/f(v)$ curve just equals T/C. When the illumination is cut off, the signal strength should diminish quickly at first and then more slowly, until it eventually drops below noise level. With a target thickness of 0·5 micron and a dielectric constant of 10 for antimony trisulphide the capacitance of a target of area 1 in^2 is about $C = 10^{-7}$ F. For a frame time $T = \frac{1}{25}$ sec the quantity $T/C = 0{\cdot}4$ V/μA. The discharge curve based on these values and on the beam acceptance curve is shown in Fig. 6, curve (*a*). The initial signal current was chosen to be 0·4 μA, the reciprocal of which is 2·5. This value was taken as the starting-point on the curve shown in Fig. 5. An area T/C was then marked out under the curve, so positioned that the ordinate 2·5 represented the average height of the area. The whole discharge curve was then constructed step by step by marking out equal areas T/C under the curve, adjacent to each other and progressing from right to left. The signals were the reciprocals of the average heights of the respective areas. As the lag arising from the discharge process is coupled with the target capacitance, it is generally called capacitance lag.

(3.2.3.2) *Photo-Conductive Lag.*

A second type of lag encountered in the tube is due to the persistence of the photo-conductive effect. When the illumina-

tion is interrupted, the photo-current does not cease immediately but decays over a period of time which may be considerable. The photo-conductive lag diminishes with increasing intensity of illumination.

The operation of the tube appears to depend to a large degree on the photo-conductive lag. This can be seen in Fig. 6, curves (*b*) and (*c*), which were actually measured in a typical VCRX360. They were obtained by scanning the tube under illumination until a signal had been built up, covering the lens and scanning the tube in the dark. The time $T = 0$ corresponds to the instant when the lens was covered up. The curve shows that the observed lag is considerably longer than that expected on the basis of pure capacitance lag. Curve (*d*) shows the lag obtained from the same tube with a much higher illumination. The reduction in lag is clearly visible.

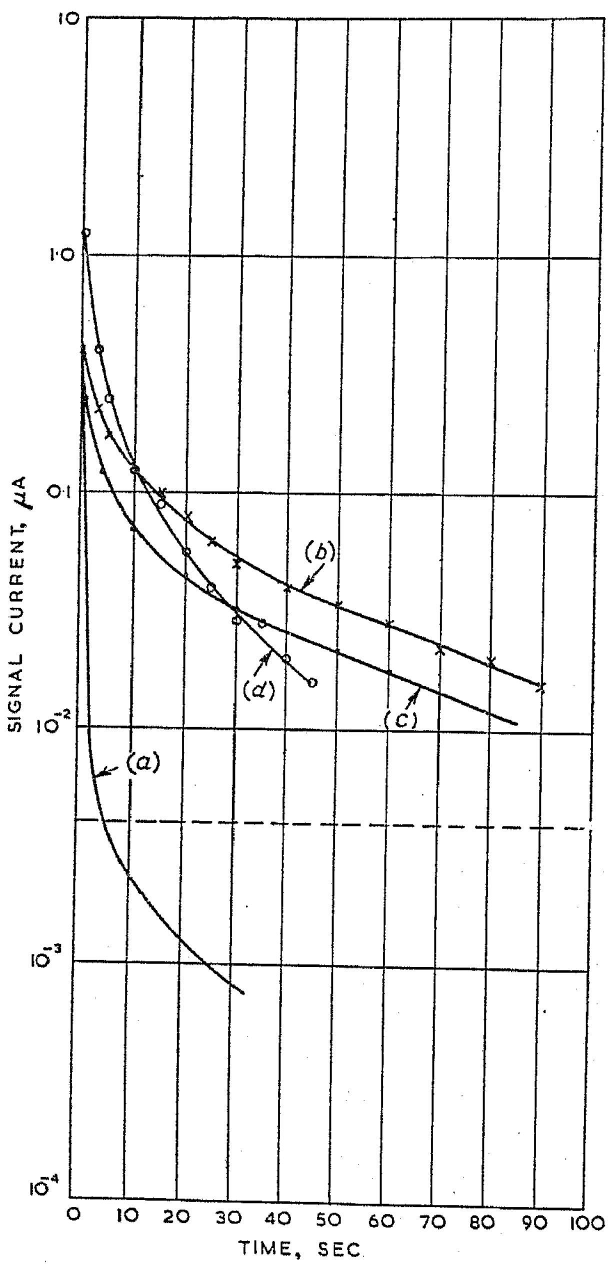


Fig. 6.—Target discharge curves of type VCRX360.
(*a*) Calculated from target capacitance on basis of measured beam-acceptance curve.
(*b*) and (*c*) Measured at low light level.
(*d*) Measured at high light level.
---- Noise level.

The effect of photo-conductive lag can also be shown by the following experiment. A signal is built up by scanning the target under illumination as described above. During the subsequent lag measurements in the dark the beam is then cut off for short periods. The signals obtained immediately after the cut-off periods are stronger than those obtained before cutting off the beam, and decay quickly to the values of the continuous scanning curve. The rise in signal is due to the persistence of the photo-current, which, in the absence of the scanning beam, causes the target potential to rise. The quick decay of the increased signals can be explained by assuming that the discharge follows the pure capacitance curve.

Confirmation was obtained by the following experiment. The target was scanned in the dark until equilibrium had been established. Then the gun cathode was switched negatively by such an amount that again the same signal was established initially and the decay of this signal was observed. The decay was much faster and closely approached the capacitance lag.

As can be seen from Fig. 6, the stored information can be read with good signal strength for well over a minute. Where enough light is available the signal/noise ratio can be made very high indeed, and ratios of over 60 dB are readily obtained. In some tubes exceptionally high reading times of 10 min and more have been observed. They are not yet fully understood but are almost certainly due to abnormal photo-conductive lag.

The lag can be reduced considerably by superimposing a steady illumination over the whole of the target. This has two effects. The target is stabilized at a more positive potential, at which the discharging power of the beam is greater.[1] It also serves to decrease the photo-conductive lag. Thus, by arranging a variable diffuse light source in the camera between the lens and the tube, the reading time can be continuously varied over a range of more than 3 : 1.

Much larger ratios have so far proved impracticable because variations both of the thickness of the target layer and of the landing velocities of the scanning beam begin to cause shading in the picture at excessive levels of illumination. Hence it was found better to control the processing of the tubes so that the correct order of storage time for any specific application was obtained, using the bias lighting merely for fine control within the limits mentioned above.

(3.2.3.3) *Build-up Lag.*

A third type of lag encountered in this tube is the build-up lag. At the onset of illumination, the discharging power of the beam is small so that the target potential rises faster than the beam can discharge it, and the signal builds up gradually. Equilibrium is reached when the target surface has risen to a potential at which the charge lost by photo-conduction is equal to the charge accepted from the beam. The higher the target capacitance the longer is the build-up lag. For the thin target layers used in the type VCRX360, the build-up lag can amount to several seconds at low levels of illumination, but it is shorter at higher intensities.

(3.2.4) Temperature Dependence.

At room temperature and in the absence of light the target layer is a very good insulator and can hold a recorded charge for about 30 min without reading. Owing to the high negative temperature coefficient of resistance, however, the holding time decreases to less than 15 min at 50° C (approximately).

(3.2.5) Light-Transfer Characteristic.

The relationship between the illumination on the antimony-trisulphide target and the output signal derived from the tube is, with few exceptions, not linear. It follows a power law of the form $S = I^{\gamma}$ where S is the signal, I the target illumination and γ is a parameter. For the antimony-trisulphide target γ is 0·6–0·7 at levels of illumination on the target up to about 0·5 ft-candle. Above this value, γ usually decreases, and at very high levels of illumination may be as low as 0·35. This means that the tube is tolerant to great overloads of light. A typical transfer characteristic is shown in Fig. 7.

(3.2.6) Sensitivity.

The operating sensitivity of the tube depends on a number of

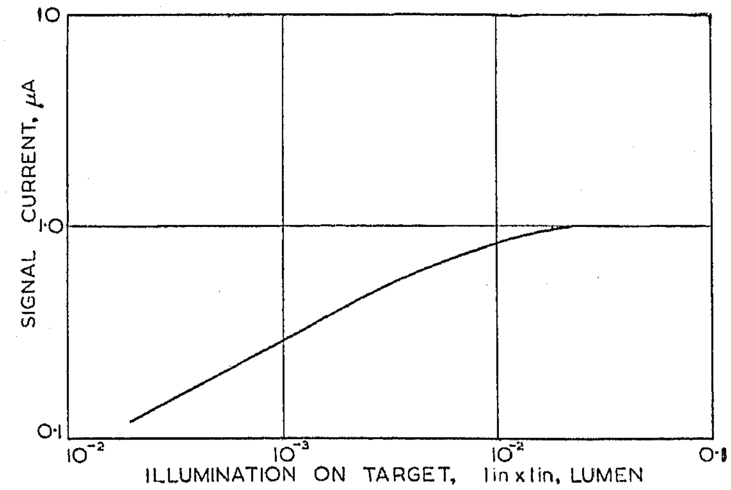


Fig. 7.—Transfer characteristic of VCRX360.

factors which make it impossible to define the sensitivity without referring to the operating conditions. First, over the normal working range, the sensitivity is proportional to the potential difference across the target layer. Moreover, owing to the fact that γ is less than unity (see Section 3.2.5) the signal rises less than proportionally with the illumination. In other words, the sensitivity in microamperes per lumen decreases with increasing illumination. Furthermore, as shown in Section 3.2.3, the signal takes a considerable number of frames to build up to its equilibrium value. Thus, with a short flash of light, as in a radar display, the full sensitivity of which the target is capable under steady illumination will not be reached. Finally, owing to the lag of the photo-conductive effect, a very short light pulse is probably less efficient than the same amount of light energy spread over a longer time. Thus the good performance obtained with the VCRX360 in a radar display can probably be attributed in part to the short afterglow, of the order of a fraction of a second, exhibited by the phosphor used in the writing cathode-ray tube.

Hence the sensitivity of the tube is referred to a steady illumination of 10^{-2} ft-candle on the target and a signal plate voltage of 10 volts. Under these conditions, the sensitivities are of the order of 900 μA/lumen while occasionally sensitivities as high as 3 000 μA/lumen have been obtained. Limits are set to the voltage across the target by electrical breakdown of the layer, which shows up as white spots in the picture, and by shading due to excessive non-uniform dark current through the layer.

(3.2.7) **Resolution.**

The high target capacitance of the tube results in small voltage excursions, and little defocusing of the beam occurs. The generation of the stored pattern does not involve the emission of electrons from the storage surface. Therefore the biasing action described in Section 3.1.2 cannot take place. Thus two limitations of the VCRX343 have been overcome, and in fact a limiting resolution of 2 000 lines on the target has been obtained. The modulation at 400 lines is 85% without aperture correction.

(3.2.8) **Colour Response.**

The colour response of the target is approximately panchromatic with a peak in the yellow-green region, the relative red response being somewhat higher than that of the eye. The best matching phosphors for the writing cathode-ray tube are a green zinc sulphide and willemite. This was found by focusing the screen of a 12-patch cathode-ray tube on the target of the VCRX360 and measuring the signals from the 12 different phosphors.

(3.2.9) **Writing Pulse.**

As the charging of the target by the light is internal to the target layer, no external write-on pulse is produced. Hence the tube can be used for simultaneous writing and reading of pulsed information such as p.p.i. displays. This is a considerable advantage over storage tubes where the writing is done by external electrons as by photo-emission or bombardment with a writing beam.

(3.2.10) **Operation in Radar System.**

As the tube was capable of very high resolution, the bandwidth of the amplifying chain was raised to $8\frac{1}{2}$ Mc/s with a linear transfer characteristic. The head amplifier used had an equivalent noise-current input of 0·004 μA. Although better figures are possible,[5] the additional complexities involved were not considered to be worth while, because the initial signal/noise ratio achieved was already very satisfactory.

The number of lines in the raster was increased to 1 001, at a field rate of 25 per second, with a 2 : 1 interlace. The final display cathode-ray tube had a willemite phosphor reducing flicker from the low field rate to negligible proportions.

The stored picture was to be a radar p.p.i. display. The circular area of the p.p.i. raster was arranged to fill the screen of the display tube completely in order to make full use of its potential resolution, the square television raster being described about this circle. The picture was kept black in the area not displayed in order to avoid troublesome flare at the edge of the screen, caused by overscanning.

In order to write the radar information, which occurs as electrical signals, into the VCRX360, a cathode-ray-tube system similar to that used for the VCRX343 was adopted. The great sensitivity of the photo-conductor reduced the demands for light output from the cathode-ray tube, thus giving a smaller spot and greater tube life. The phosphor used was willemite. Light bias was provided, when required, by four 2-watt bulbs situated between the lens and the VCRX360. The lens used was of 2 in focal length and was operated at an aperture of f2. This lens was specially selected after tests had shown that one of normal good quality had caused a noticeable degradation of picture quality.

(3.2.11) **System Performance.**

(*a*) *Storage Time.*—Storage times were adequate, the maximum obtainable varying somewhat from tube to tube in the range 1–10 min. As shorter storage times are also required, some tubes have been made with maximum storage times of about 10 sec. Light bias had the effect mentioned in Section 3.2.3.

(*b*) *Resolution.*—The storage tube itself is capable of resolving at least 1 600, and usually 2 000, lines in the centre of the picture. The cathode-ray tube used can resolve at least 1 200 lines, and, in the centre, over 1 600 lines. The lens used is capable of resolving at least 3 000 lines over the whole field. It was therefore to be expected that the combined cathode-ray-tube and storage-tube system would be capable of resolving over 1 000 lines in the centre of the picture. When making resolution tests, the reading scan was reduced in size, thus reducing the speed of the scanning spot and hence eliminating any limitations introduced by amplifier bandwidth and display cathode-ray-tube spot size.

By visual estimation a limiting resolution for the system of approximately 700 lines was actually achieved over most of the target area. The apparent discrepancy may be due to either or both of two causes. First, the focusing procedure is rather intricate, three separate items requiring adjustment. Secondly, the overall test was made using fine single spots, whereas the tests on individual components were made using bar systems, which give the eye a much better chance to resolve fine structure.

(*c*) *General.*—The overall appearance of the picture is quite clean, shading not being obtrusive, and a high proportion even of experimentally produced tubes show only a few spots. The

storage time gives a lag roughly equal to that of the average fluoride cathode-ray tube; but using a 16 kV white or green tube in the final display, the picture is visible in a room quite brightly lit, even with the tube face towards the window. The 1 001-line raster, although noticeable, is not obnoxious, and does not appear to degrade the picture.

It is believed that the main agent in the limitation of resolution is the writing cathode-ray tube. Efforts are currently in hand to improve the resolution of the system so that the overall results are not inferior to those of a simple directly-viewed cathode-ray tube. These include using a larger storage-surface area and raising the final anode potential on the writing-on cathode-ray tube.

The transfer characteristic of the complete system is somewhat complex, as the writing cathode-ray tube has a value of γ of about 2, and the storage tube one of about 0·6. Therefore, unless black levels are held to a fixed value, and only one writing-on operation is performed at each spot, the overall value of γ can vary quite widely according to the nature of the method of measurement. If black is equivalent to zero signal, the overall response is approximately linear.

(4) TUBES OPERATING WITH BEAM MODULATION

(4.1) General

An approach alternative to the charge-restoration principle was to utilize a small charge to control an electron beam without the charge itself being destroyed in the process. A variety of television camera tubes had already been proposed using this principle to enhance light sensitivity.[10,11] However, in modern television systems the sensitivities of camera tubes are beginning to approach the limit set by available photo-cathode materials, and it seems unlikely that the idea will now be pursued in this field.

With storage tubes, however, the beam-modulation principle is finding increasing application, and descriptions of such tubes have already been published by various authors.[3,4,12,13,14]

(4.2) Beam-Modulation Tube Type VCRX326

(4.2.1) Description.

A diagram of the VCRX326 is shown in Fig. 8. It contains three plane-parallel electrodes, a continuous photo-cathode 1, an accelerating mesh 2, and a storage target 3. An enlarged sketch of the target is also shown in the Figure. A metal

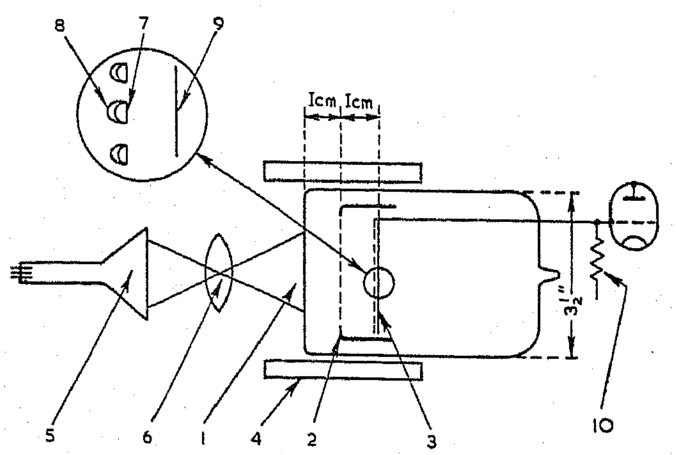


Fig. 8.—Beam-modulation tube VCRX326.

1. Photo-emissive cathode.
2. Accelerating mesh.
3. Storage target.
4. Focus coil.
5. Writing and reading cathode-ray tube.
6. Lens.
7. Storage mesh.
8. Insulating coating.
9. Backing plate.
10. Signal resistor.

mesh, 7, is coated on the side facing the photo-cathode with an insulating layer of calcium fluoride, 8. At a distance of 1 mm from the mesh is a metal backing plate, 9. An axial magnetic field of 300 oersteds is provided by the solenoid, 4, which focuses electrons leaving the photo-cathode on to the storage target. These photo-electrons perform the functions of writing, reading and erasing according to the potentials applied to the three electrodes. For reading, the photo-emissive cathode is illuminated with a small spot of light which is scanned over its surface. This is obtained by optical projection of the fluorescent spot of the cathode-ray tube, 5, with the lens, 6. The resulting beam of photo-electrons in the storage tube may also be used for erasing and for writing. An alternative method of writing is by optical projection on to the photo-cathode of an image of the picture to be stored. Similarly, erasure may be carried out by a pulsed light source. Since the tube is essentially a single-beam tube, only one function is possible at a time. However, as described later, reading and erasure may readily be alternated at a sufficiently high rate to be imperceptible.

(4.2.2) Reading and Writing.

In the reading operation, electrons arriving at the target stabilize the insulator to a potential approximately equal to that of the photo-cathode (zero voltage). The mesh, 7, and backing plate, 9, are connected together and operated at a small positive potential of about +3 volts. In this condition a large fraction of the beam electrons are drawn through the mesh 7, and a steady current flows in the resistor 10.

The picture to be "written on" is applied as an optical image to the cathode, the potential of which is lowered to about −3 volts. The potential of the surface of the insulator then changes towards this value, the rate of change being determined by the intensity of illumination. By adjustment of the exposure, a charge image is left upon the insulator in which the potential range varies from zero to −3 volts.

The cathode is returned to zero voltage, and in the subsequent reading operation, the current flowing in the resistor 10 is modulated by the charge pattern. High lights in the applied picture correspond to an insulator potential of −3 volts, which is sufficiently negative to reduce the signal current to zero. Low lights will not have affected the insulator potential, and here the original signal current flows. At intermediate potentials, signal currents between the limits stated will obtain. The range of insulator potentials is thus negative to the photo-cathode during the reading period. Accordingly, electrons originating at the cathode cannot land upon the insulator.

The above method of operation produces maximum signal current in the low lights of the picture. Thus fluctuation noise current is also a maximum. With a reading raster operated at television frequencies, this is not of great importance, as the noise level of the amplifier channel is greater than the fluctuation noise of the signal. For some applications it may perhaps be undesirable, and an alternative method which inverts the polarity of the signals may be adopted.[14]

(4.2.3) Erasure.

Controlled erasure requires that the insulator potential be re-stabilized to its datum potential of zero voltage at any desired rate. Re-stabilization is a twofold operation. First, the photo-cathode potential is made sufficiently negative so that more secondary electrons are released from the insulator than primary electrons arriving. With illumination incident on the photo-cathode, the insulator potential then drifts positively until it is arrested by coplanar biasing of the exposed metal mesh. Secondly, the photo-cathode potential is returned to zero voltage, and the positive insulator discharges to the datum potential.

When the reading raster is of the television type, the erasure and reading operations may be conveniently alternated. During the line-flyback periods of the reading raster, a pulse of about -150 volts is applied to the photo-cathode and a variable pulse to the modulator of the cathode-ray tube. Thus, during these periods erasure conditions obtain, and control of the cathode-ray-tube brightness determines the erasure rate.

(4.2.4) **Positive-Ion Currents.**

Positive ions are formed on both sides of the accelerating mesh during writing, reading and erasure. Most of those formed on the target side of the mesh land on the insulator and charge it positively. This action is essentially the same as that of the beam when erasing a charge pattern, and the ion current thus limits the duration for which reading may be continued.

The probability of an electron ionizing a gas molecule in its transit across the tube is a function of the distance the electron travels and the potential of the accelerating mesh. To minimize the probability it is desirable to make these quantities small. A distance of 2 cm has been chosen in this tube as a matter of practical convenience. On the other hand, it is found that the resolution of the tube decreases as the accelerating-mesh potential is lowered. A compromise must therefore be chosen between maximum reading time and resolution according to the application.

By lowering the mesh potential to $+3$ volts, the probability of ionization is reduced to zero.[15] In this condition the maximum reading time is determined only by the natural leakage of the insulator. It has been found that this can be several days.

(4.2.5) **Writing Speed.**

Writing speed is a function of the beam current leaving the photo-cathode and of the target capacitance. The beam current is set by the amount of light incident on the cathode and the photo-sensitivity. Using a 20 kV cathode-ray tube with a zinc-oxide phosphor as a source of light, an antimony-caesium photo-cathode and a wide-aperture lens, a beam current of about $0 \cdot 5\,\mu$A may be obtained. This is sufficient to charge the insulator to its cut-off voltage in $\frac{1}{25}$ sec.

(4.2.6) **Performance.**

Accelerating-mesh potential	200 volts	100 volts	3 volts
Limiting resolution (writing by optical projection of test chart on to photo-cathode)	800 lines	500 lines	400 lines
Maximum reading time	15 min	45 min	60 hours
Minimum erasure time (with alternated reading)	About 5 sec		
Initial signal/noise ratio	30 dB	30 dB	30 dB

(4.3) Beam-Modulation Tube Type VCRX350

(4.3.1) **Description.**

By the inclusion of separate writing and reading guns in the VCRX350, a disadvantage of the VCRX326—namely that writing and reading were necessarily sequential—has been overcome. The two guns, as shown in Fig. 9, are situated in line on opposite sides of the storage target. This arrangement obviates the disadvantages of oblique scanning and focusing, which are encountered in some present-day storage tubes.[3,16,17]

The storage target, 1, consists of a metal mesh coated on the side facing the gun, 8, with an insulating layer. This surface is scanned by the modulated writing beam, deflected into a suitable raster. Secondary electrons are then released from the insulator in excess of the primary electrons arriving, and a positive charge pattern results. The potential of this charge is limited by the field of the coplanar metal mesh, resulting in stability for repetitive writing.

The reading section has been developed from the C.P.S.

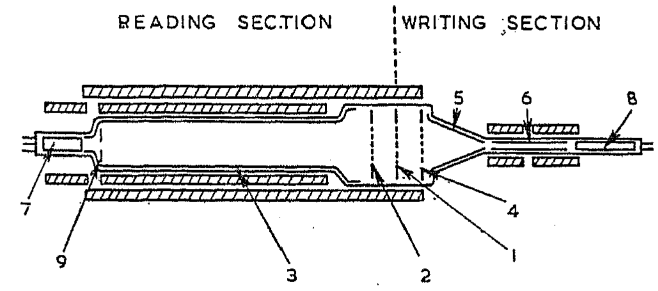


Fig. 9.—Beam-modulation tube VCRX350.

1. Storage mesh, 0 to +4 volts.
2. Ion-trap mesh, 10 volts positive to wall anode 3.
3. Wall anode, +400 to +500 volts.
4. Erasure mesh, −5 to +1 000 volts.
5. Wall anode, +50 volts.
6. Wall anode, +50 volts.
7. Reading gun, cathode zero voltage, anode +500 to +600 volts.
8. Writing gun, cathode −1·5 to −4 kV, anode +50 volts.
9. Writing-beam interceptor.

Emitron.[2] The electron beam arrives at the storage mesh at a uniform velocity which is nearly zero. Under the action of the positive charge pattern on the insulator, a variable fraction of the beam then passes through the interstices of the mesh and is collected on the erasure mesh, 4, and the wall coatings, 5 and 6, of the writing section. Most of the remaining beam electrons return to the reading gun, 7, from which signals are taken. The remainder are intercepted by the ion-trap mesh, 2.

The potential of the erasure mesh, 4, controls the rate of decay of the stored charge, and its action is described in Section 4.3.3.

The datum (i.e. zero signal) potential of the insulator is approximately that of the reading-gun cathode. Although the charge pattern is positive with respect to this cathode when reading, the insulator is shielded from the direct discharging action of the beam by the physical barrier of the metal storage mesh.

(4.3.2) **Storage Target.**

The behaviour of the tube is a function of various storage-target parameters described below. In order to find the optimum conditions, a series of tubes was made in each of which a number of storage targets varying in one parameter only were mounted side by side.

Over the range of mesh transparencies available (30–75%) the output signal/noise ratio increased approximately as the transparency. A similar variation was found in the minimum erasure rate. However, since the high-transparency meshes were very fragile, a value of 60% was adopted.

The mesh pitch was found to have little effect upon the maximum reading time, the signal/noise ratio and the minimum erasure time. To obtain good resolution, a fine-pitch mesh is desirable. Manufacturing difficulties increase as the pitch of a mesh is reduced, however, particularly if a mesh of high transparency is needed, and for these reasons a pitch of $0 \cdot 0016$ in was chosen.

The bars of the meshes used have a roughly triangular cross-section. On one surface of the mesh the bases of the triangles are exposed, and on the other the vertices. With the insulator applied to the surface containing the vertices, superior signal/noise ratio, quicker erasure rate, and a longer maximum reading time were obtained.

The insulator thickness determines the capacitance of the storage target, and, in general, the results obtained by varying the thickness may be explained solely in terms of variation of the capacitance. Thus, as the insulator thickness is reduced the writing speed is also reduced, and the minimum erasure time and maximum reading time are increased. Using calcium fluoride, a thickness of $0 \cdot 5$ micron was found to give a writing speed of about $0 \cdot 3$ microsec per picture element, a minimum

erasure time of about 3 sec, and a maximum reading time of 30 min for a high-definition picture.

(4.3.3) **Erasure.**

Erasure of the stored charge is carried out by the reading beam.[18] When the charge pattern is being read, a number of the electrons passing through the storage mesh are reflected on to the insulator, and they discharge it to its datum potential. The number reflected is primarily determined by the potential of the erasure mesh, 4. With small negative potentials, the quickest erasure time of about 3 sec is obtained. As the mesh is made progressively more positive the erasure time increases, until at a potential of about +1 000 volts, continuous reading of a high-definition picture for more than 30 min is possible.

The erasure rate is also dependent upon the potential of the storage mesh and the quantity of stored charge. With the storage mesh at the potential of the reading-gun cathode, or slightly negative to it, the quickest erasure rate is found. Furthermore, small-amplitude charges decay more rapidly than those of large amplitude, which is advantageous when the applied information contains an appreciable quantity of random fluctuations or noise. However, although erasure may be complete after one write-read-erasure operation, so far as the reproduced picture is concerned, a residual charge still remains on the insulator. This may be examined by raising the potential of the storage mesh by about $\frac{1}{2}$ volt, when a weak signal compared with the original is obtained. The importance of the weak residual charge depends upon the application. When a relatively large amount of noise is being written on to the target, the weak charge is masked by this noise and is of no account. When the information to be stored has a high signal/noise ratio, it is necessary to re-stabilize the insulator to its datum potential by depositing a uniform charge and erasing. This operation is required after each write-read cycle.

With more positive potentials on the storage mesh, up to about 4 volts, the minimum erasure time and also the maximum continuous reading time are increased.[18] The residual charge mentioned above is appreciably reduced, and a storage-mesh potential in this range is preferable for a low-noise application. Even so, it may be desirable to re-stabilize the insulator after each write-read cycle.

With even more positive potentials on the storage mesh, shading is introduced into the reproduced picture by non-uniform landing of the reading beam on the exposed metal of the storage mesh. The modulation for a given stored charge also decreases, and at a mesh potential of about +10 volts the phase of the reproduced signal is inverted. In this condition, even with the erasure mesh at a negative potential, the minimum erasure time is increased enormously and may be 30 min or more. The reasons for these latter effects are not understood.

The erasure rate is also dependent upon the area of the stored charge. In general, the smaller the diameter of the charge or the higher the definition, the more difficult it is to hold the picture. With a potential of +4 volts on the storage mesh and 1 000 volts on the erasure mesh, the maximum reading time for a spot of one five-hundredth of the picture diameter is about 30 min. The initial signal/noise ratio is then 25 dB. If the writing beam is defocused so that the spot diameter is one two-hundredth of the picture diameter or greater, the maximum reading time increases to 2–3 hours. It is believed that this effect is due to the coplanar biasing action of the areas surrounding the charged spot.

The uniformity of erasure has been of most interest in the latively short reading times required for the radar application. hese times, no appreciable variation of erasure rate has been ved over the picture area.

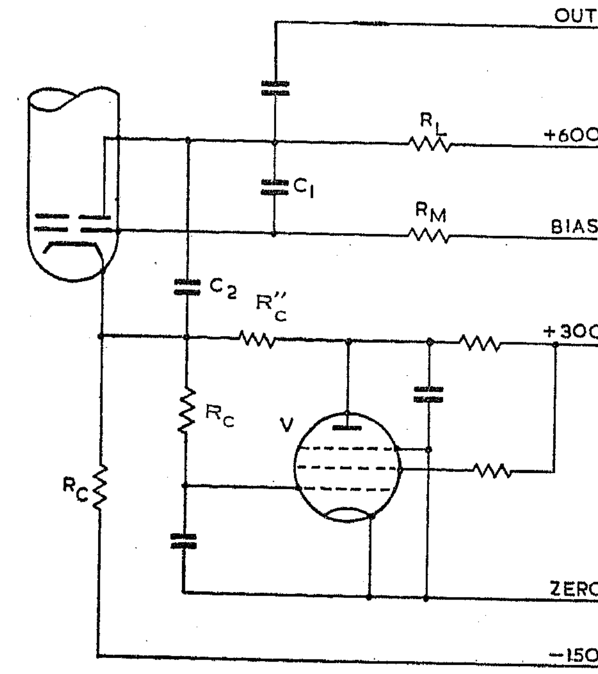


Fig. 10.—VCRX350 reading-gun circuit.

The valve V stabilizes the cathode potential of the reading gun against low-frequency drift. Its response to the high frequencies of the picture signals is negligible.

(4.3.4) **Reading Gun.**

The reading gun is used as the output-signal electrode of the tube and is connected according to the circuit[19] of Fig. 10. The three electrodes, namely cathode, modulator and limiter, have steady voltages applied to them through high resistances R_C, R'_C, R''_C, R_M, and R_L. The capacitances C_1 and C_2 are of low impedance at video frequencies compared with R_C, R'_C, R''_C, R_M, and R_L. Thus for these frequencies the gun may be considered as a single electrode. The input impedance of the amplifier is also effectively low compared with the above resistances, and the bulk of the signal current flows therein. The valve V stabilizes the mean potential of the cathode.

The above arrangement allows signals to be taken from the gun end of the tube without increasing the fluctuation noise in an $8\frac{1}{2}$ Mc/s amplifier by more than 1·0 dB. This is so because the fluctuations in the cathode current of the gun do not flow in the signal resistor, but return to cathode via C_1 and C_2. The sum of the fluctuations contained in the outgoing and returning electron beams remains. With an ion-trap mesh transparency of 70%, about half the outgoing beam returns to the gun in the absence of a stored picture. For equal outgoing and returning electron-beam currents, it has been shown* that the total noise is given by

$$i_N^2 = 4ei_B\left[f - \frac{\sin 2\pi f\tau}{2\pi\tau}\right] \quad . \quad . \quad . \quad . \quad (4$$

where i_N = Total noise current.
e = Electron charge.
i_B = Electron-beam current.
f = Pass band of amplifier.
τ = Transit time of electron to target and back.

To allow for the 50% absorption on the ion-trap mesh, it is assumed that 50% of the outgoing beam is subject to the above formula, and the other 50% generates noise according to the equation

$$i_N^2 = 2ei_Bf \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

This ignores any partition noise arising from the interception of current by the ion-trap mesh.

For an outgoing beam current of 1 μA with $\tau = 5 \times 10^{-8}$ sec and $f = 8\frac{1}{2}$ Mc/s, the total beam noise computed is

$$i_N = 1\cdot89 \times 10^{-9}\,\text{amp}$$

* The authors are indebted to Dr. B. Meltzer for this calculation.

For the same frequency range the noise current of a typical amplifier is 4×10^{-9} amp. The total noise is thus increased by 0·88 dB to $4 \cdot 42 \times 10^{-9}$ amp.

(4.3.5) **Writing Gun and Writing Speed.**

The writing gun is of triode construction. It is focused and deflected magnetically, and at an anode-to-cathode potential of 2000 volts it delivers a beam current of 1–2 μA. The diameter of a reproduced spot may be measured on a monitor cathode-ray tube by reducing the reading-scan amplitudes, and thus avoiding any bandwidth limitations. The figure obtained in this manner is 0·096 mm, and hence both the writing and reading spots must be smaller than this.

No effort has been made to increase the writing speed beyond that required for the radar application. With the present design, used in a radar display, reproduced spots of one five-hundredth of the picture diameter have been written in 5 microsec with an output signal/noise ratio of 25 dB.

Use as a store for a television picture (see Section 4.3.13) indicated that writing speeds of 0·3 microsec per spot diameter could be achieved with an output signal/noise ratio of 20 dB.

(4.3.6) **Spurious Signals.**

A fraction of the writing beam travels the full length of the tube and is scanned over the reading-gun anode and the wall anode by the combined reading and writing deflection fields. Spurious signals of opposite phase to the reading signals result, but their time duration and amplitude have been much reduced by the introduction of the apertured plate, 9, of Fig. 9. The aperture ($\frac{3}{8}$ in) is sufficiently large not to intercept any of the reading-beam electrons.

(4.3.7) **Positive-Ion Currents.**

With the electrode potentials normally used, only the positive ions formed between the ion-trap mesh and the erasure mesh are absorbed on the storage target. The spacing between these two meshes is small (2 cm) as is the ion current. Using an average tube, the shading or change in the background level due to ion deposition becomes visible with continuous reading of any one picture only after some 2–3 hours.

(4.3.8) **Shelf-Life of a Stored Picture.**

Re-examination of a stored picture, with little loss of signal strength, may normally be made several days after it is first stored, provided that the tube is switched off in the interim. In one instance a picture was detectable after two months. The picture was then easily erased and normal operation of the tube was resumed, showing that there was no burn-in of the stored image.

(4.3.9) **Transfer Characteristic.**

A curve of the transfer characteristic is shown in Fig. 11. It is seen that a useful linear range of 25–30 dB exists between the output signal and the primary charge leaving the writing gun.

(4.3.10) **Scanning.**

The equipment used for the VCRX360 was designed to accommodate both this tube and the VCRX350. The reading-section scan and focus coils require almost identical amounts of drive, and the writing section, owing to a lower final-anode potential (1·5 kV instead of 25 kV) and smaller scan angle (28° instead of 38°), requires a smaller amount of scan drive than the VCRX360 system.

The VCRX350 is somewhat susceptible to the effects of external magnetic fields, and in order to obtain good resolution, careful magnetic screening of the tube is necessary.

(4.3.11) **Black Level.**

The advantages of establishing black level at the end of each

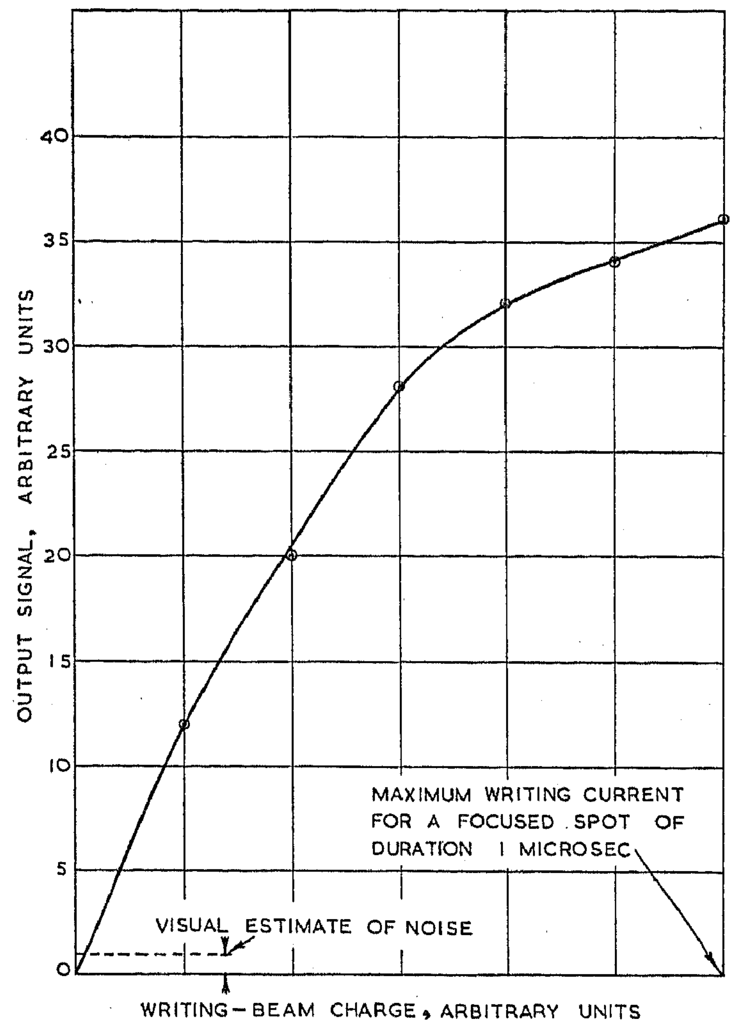


Fig. 11.—VCRX350 transfer curve.

television scanning line are well known. In order to obtain the reference black level, the reading beam is blanked out during flyback by lowering the potential of the storage mesh by about 10 volts. However, when the scanned area is described about the circle of the storage mesh, which is desirable in order to obtain maximum resolution from the tube, the reading beam is intercepted at the corners of the raster by the reading-section wall anode. This gives rise to a large white signal, which exists along the whole of the top and bottom lines.

Therefore, the reading-scan raster is reduced to 90% of its former linear dimensions, and under these circumstances at least the middle half of every return line is available as true reference black.

Non-uniform transmission by the ion-trap mesh still renders this signal liable to variations from point to point in the picture, but by careful control in tube manufacture these may be kept down to less than 5% of peak white.

Because of non-uniform landing velocities induced by scanning, the black level tends to be lower at the edges of the picture than in the centre. Careful scanning and focus-coil design enables this shading to be reduced to a tolerable value.

(4.3.12) **Resolution.**

Using the television standards defined in Section 3.2.9, the smallest reproduced spot which is regularly achieved is one five-hundredth of the diameter of the picture, although on occasions this figure has been improved to one seven-hundredth. These figures refer to spots in the centre of the picture, the spot size increasing somewhat towards the edge. Spot sizes of less than one five-hundredth of the picture diameter have, however, occasionally been achieved at the picture edge.

(4.3.13) **Television Application.**

The experiment was made of writing a stationary British Standard television picture of 1, 2, 3 or more frames. It was

Table 1

System using tube type	System resolution	Writing speed	Storage time	Background quality	Geometry	Initial signal/noise ratio	Decay law	Simultaneous write/read	Input/output coupling
	No. of spots per picture diameter	μs per spot diameter				dB			
VCRX 343	400	5	8 sec (max.)	Good	Good	12*	Near-exponential fair	Yes	Bad
VCRX 360	700	0·1	3 sec (min.) 10 min (max.)	Good	Good	26* 36†	Poor	Yes	None
VCRX 326	550**	0·3	5 sec (min.) 15 min (max.)	Fairly good	Excellent	30*†	Not measured probably linear	No	—
VCRX 350	700	0·3	3 sec (min.) 30 min (max.)	Fair	Good	20* 30†	Good	Yes	Small

* At stated writing speed. † Best obtainable. ** Estimated.

found that three frames were necessary to obtain an output signal/noise ratio of 30 dB. However, an acceptable picture of 20 dB signal/noise ratio, with good half-tone rendition, was obtained by writing a single frame.

(4.3.14) **Possible Developments.**

The use of a rectangular-mesh electrode structure with standard 3 : 4 aspect ratio and height equal to the diameter of the present mesh would eliminate the black-level troubles described in Section 4.3.11. It would also enable the whole of the storage surface to be used when storing one frame of a television picture.

It has been mentioned in Section 3.2.10 that flare occurs around the edge of the 12 in picture tube because of overscanning. This is particularly troublesome with the VCRX350 because there is a bright white surround to the picture, but this would be reduced in the modified tube proposed.

It would be eliminated entirely if no signal were written in to the tube in the areas corresponding to those outside the area of the screen of the viewed cathode-ray tube. This could be done either by scan limitation of the writing raster external to the tube, or by the use of a circular shield mesh screening the storage mesh in those areas.

(5) SUMMARY OF SYSTEM PERFORMANCE (see TABLE 1)

Resolution.—In all cases the resolutions quoted have been those obtained on a complete system including the final display cathode-ray tube.

Writing Speed.—No measurements have been made to obtain the maximum writing speed obtainable from the tubes. The figures quoted have been obtained readily in practice. For the VCRX350 the quoted speed gave an initial signal/noise ratio of 20 dB (see Section 4.3.13), and for the VCRX360, an initial signal/noise ratio of 26 dB.

Storage Time.—The times given are for continuous reading, which is assumed to have terminated when the signal has disappeared into noise. The stored signal in each case was a spot of a size approaching the limit of the resolving capabilities of the tube (see Section 4.3.3 for storage time for larger areas).

Background.—The background in all tubes is assumed to be that which would be obtained in large-scale production, and under these circumstances, considerable reduction of dust spots, etc., can be expected compared with present-day experimental tubes.

Geometry.—It was not the purpose of the present investigation to produce a tube of sufficient accuracy to enable measurements to be made of target position on the final display. All that was required was that the final display should not be obviously different from that obtained by direct presentation of a radar picture.

Initial Signal/Noise Ratio.—The figures quoted are those which were obtained when using the head amplifiers described in the text. That used for the VCRX343 was also used for the VCRX326.

Decay Law.—The ideal decay law for a radar application is probably linear, or some law not very far removed from this. Fig. 12 shows typical decay laws for three tubes normalized to equal storage times.

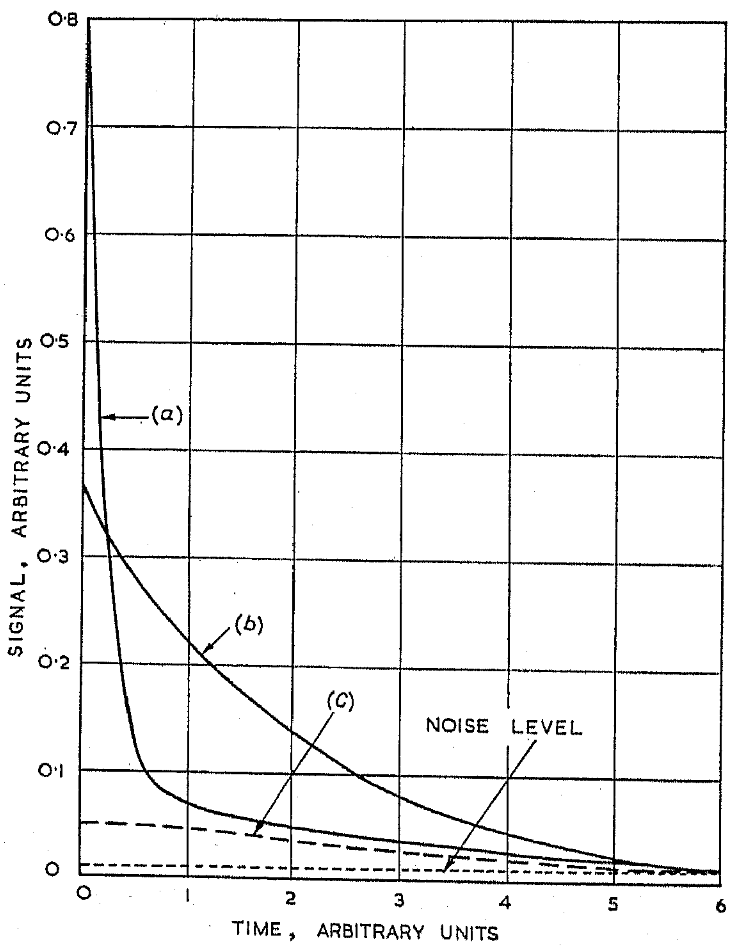


Fig. 12.—Decay characteristics for tubes type VCRX343, VCRX350 and VCRX360 normalized to equal storage times.

(a) VCRX360.
(b) VCRX350.
(c) VCRX343.

Simultaneous Write-Read.—This item is self-explanatory.

Input/Output Coupling.—With the VCRX343 the input/output coupling is inherent in the tube, but can be reduced to a very small amount. In the VCRX350 it has been minimized by careful design. In the VCRX360 it does not exist, and in the VCRX326 it does not arise because of the inability of this tube to write and read simultaneously.

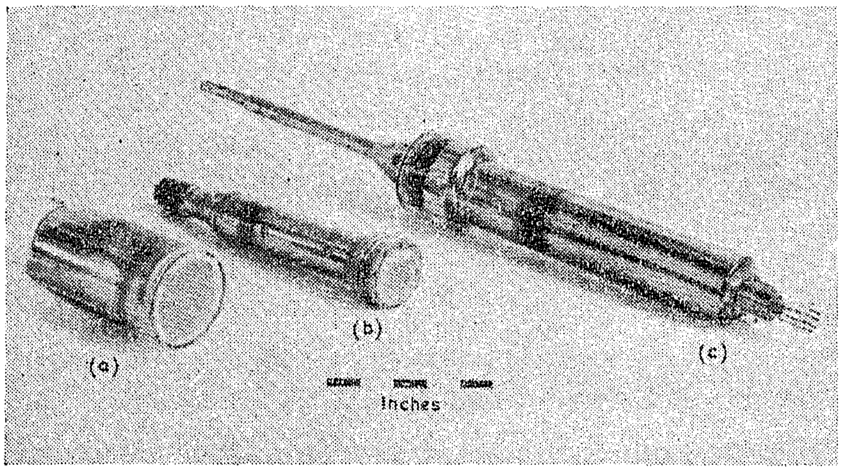


Fig. 13.—Photographs of the tubes.
(a) VCRX326.
(b) VCRX360.
(c) VCRX350.

(6) ACKNOWLEDGMENTS

Thanks are due to Prof. J. D. McGee for many helpful suggestions made during the course of the work, to Mr. H. E. Holman, who supplied the meshes used in the tubes, to Dr. B. Meltzer, who carried out calculations, and also to Messrs. A. C. Dawe, P. L. Holmes, W. J. C. Hosking, P. Muff, S. Taylor, J. Wardley, and the others too numerous to mention—without whose help the programme could not have been carried out. Thanks are also due to the Admiralty under whose auspices much of the work was carried out, and to Mr. G. E. Condliffe, Managing Director of E.M.I. Research Laboratories Ltd., for permission to publish the paper.

(7) REFERENCES

(1) FORGUE, S. V.: *R.C.A. Review*, 1947, **8**, p. 633.
(2) MCGEE, J. D.: "A Review of some Television Pick-Up Tubes," *Proceedings I.E.E.*, Paper No. 931 R, April, 1950 (**97**, Part III, p. 377).
(3) KNOLL, M., and KAZAN, B.: "Storage Tubes" (John Wiley, 1952).
(4) HAEFF, A. V.: *Electronics*, September, 1947, **20**, p. 80.
(5) JAMES, I. J. P.: "Fluctuation Noise in Television-Camera Head Amplifiers," *Proceedings I.E.E.*, Paper No. 1300 R, April, 1952 (**99**, Part IIIA, p. 796).
(6) CASSMAN, H.: British Patent No. 657634, 1946.
(7) LUBSZYNSKI, H. G.: British Patent Application No. 28591, 1954.
(8) JAMES, I. J. P., and STILLWELL, P. F. T. C.: British Patent No. 711275, 1950.
(9) LUBSZYNSKI, H. G.: British Patent No. 665877, 1948.
(10) MCGEE, J. D., and LUBSZYNSKI, H. G.: British Patent No. 510696, 1937.
(11) STEWART BROWN, W.: British Patent No. 539419, 1940.
(12) HERGENROTHER, R. C., and GARDNER, B. C.: *Proceedings of the Institution of Radio Engineers*, 1950, **38**, p. 740.
(13) SMITH, S. T., and BROWN, H. E.: *ibid.*, 1953, **41**, p. 1167.
(14) KNOLL, M., RUDNICK, P., and HOOK, H.: *R.C.A. Review*, 1953, **14**, p. 492.
(15) WEBLEY, R. S.: British Patent No. 715681, 1950.
(16) PENSAK, L.: *R.C.A. Review*, 1949, **10**, p. 59.
(17) PENSAK, L.: *R.C.A. Review*, 1954, **15**, p. 145.
(18) WEBLEY, R. S.: British Patent Application No. 11478, 1953.
(19) WEBLEY, R. S.: British Patent Application No. 30716, 1953.

DISCUSSION ON "AN ANALOGUE COMPUTER FOR USE IN THE DESIGN OF SERVO SYSTEMS"*

SOUTH-EAST SCOTLAND SUB-CENTRE, AT EDINBURGH, 19TH OCTOBER, 1954

Mr. J. B. Smith: The paper was written at a time when little had been published on systematic servo-mechanism design, and the computer fulfilled a valuable function in assisting designers to understand the principles involved. During the past four years much work on the subject has been published, and most designers have adopted one or other of the various techniques utilizing specialized design charts and/or function tables. Does the author feel that his computer can still be used to advantage in some aspects of design?

Mr. B. B. English: While the design of servo systems has been advanced as the primary purpose of the computer, it has also been suggested that it may be of use in teaching servo-mechanism theory. However, I find it difficult to believe that a computer of this type would be as convincing and as easy to understand as a purely electronic simulator. Most students of servo mechanisms will have had a basic training in circuit theory and electronics, and will be able to grasp the mode of operation of a simulator with its direct comparison with the actual servo mechanism much better than the purely mathematical approach of the computer described in the paper.

Has the author used the computer for teaching purposes, and has he found its mode of operation easy to explain?

Dr. E. E. Ward (*in reply*): In reply to Mr. Smith, my view is that the question least understood in current servo-system theory is the relation of the required output function—particularly when stated as a time function—to the specifications of the apparatus forming the open loop, particularly when regarded as a frequency function. A computer of the kind described should be useful in making such calculations. In reply to Mr. English, I have not had occasion to use the computer for teaching purposes.

* WARD, E. E.: Paper No. 1201 M, November, 1951 (see 99, Part II, p. 521).

621.374.32 : 621.314.7 : 681.142

Paper No. 1858 R
July 1955

A TRANSISTOR DIGITAL FAST MULTIPLIER WITH MAGNETOSTRICTIVE STORAGE

By G. B. B. CHAPLIN, M.Sc., Ph.D., Graduate, R. E. HAYES, B.Sc., Student, and A. R. OWENS, B.Sc., Graduate.

(*The paper was first received* 16*th December*, 1954, *and in revised form* 9*th March*, 1955.)

SUMMARY

The transistor is gradually establishing itself as a useful circuit element, but although individual circuits have proved to be adequately reliable their collective reliability in a more complex apparatus has not hitherto been proved. Accordingly a digital fast multiplier—an important part of any large-scale computer—has been constructed using transistors throughout.

Operating at a basic frequency of 125 kc/s, the multiplier computes the product of two 32-digit serial numbers in about 4 millisec.

The logical design has a novel feature in that the digits of the multiplier are taken in pairs instead of singly.

The multiplier can be divided into three parts: the generator of timing waveforms, the arithmetic unit and the storage system. Point-contact transistors are used throughout except in the amplifiers associated with the storage system, where junction units are used. The actual storage elements are magnetostrictive delay lines; some of the results of an investigation into their operation are included as appendices.

Out of a total of 84 point-contact transistors, 12 junction transistors and 568 crystal diodes, 3 point-contact units have failed in 7 months of daily operation. No selection of transistors is required. The total power consumption is 50 watts.

It is concluded that transistors of the types at present available are well suited to computer circuits and that both the construction and maintenance of the apparatus are facilitated by their use.

LIST OF PRINCIPAL SYMBOLS

v_e, v_b, v_c = Electrode voltages.
i_e, i_b, i_c = Electrode currents.
I_e, I_b, I_c = Supply currents.
i_{c0} = Collector current with zero emitter current.
α_k = Ratio i_c/i_e at knee of i_c/v_c characteristic.
x = Multiplicand (number).
y = Multiplier (number).
$\left.\begin{matrix} y_A \\ y_B \\ y_C \\ y_D \end{matrix}\right\}$ = Sections of y.
i_D = Diode current.
T = Transistor.
D = Diode.
S = Sending circuit.
$C(x)$ = Complement of x.
l = Length of delay-line coil.
c = Velocity of sound in nickel.
X = Waveform defining x period.
Y = Waveform defining y period.
Y_1, Y_2, Y_3, Y_4 = Waveforms defining periods of y_A, y_B, y_C, y_D.
$\left.\begin{matrix} \text{PS} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \\ \text{F} \end{matrix}\right\}$ = Other timing waveforms.
$\left.\begin{matrix} \text{PS}' \\ \text{A}' \\ \text{B}', \text{etc.} \end{matrix}\right\}$ = Self-resetting inversions of PS, A, B, etc.
$\left.\begin{matrix} \text{Clock = Clock waveform} \\ \text{I-clock = Inverse clock waveform} \\ \text{Strobe = Strobe waveform} \\ \text{I-strobe = Inverse strobe waveform} \\ \text{D-strobe = Delayed strobe waveform} \end{matrix}\right\}$ Basic waveforms.

(1) INTRODUCTION

The transistor is gradually establishing itself as a useful circuit element, and although it is too early yet to assess its reliability in general terms, its short-term reliability appears promising.

Whereas the junction transistor seems to be applicable mainly to linear amplifiers, the point-contact unit is shown to best advantage in the field of 2-state circuits. Such a field is typified by the electronic digital computer, which should benefit from the small size of the transistor, the elimination of heaters and the resulting simplified construction. Although many individual computing circuits have been devised,[1,5,10] proof of the ability of these circuits to operate reliably together in a computer has hitherto been lacking. The apparatus described in the paper was devised to provide this proof.

An important part of all large-scale computers is the multiplying system, and for reasons of economy it was decided not to build a complete computer but to construct a digital fast multiplier, including a storage system and units for the generation of the numbers and all necessary waveforms. The logical design is based on the multiplier used by Dr. Kilburn in a thermionic-valve computer in the University of Manchester,[4] and the project was carried out under his general direction. The system, in which the digits are taken in pairs, lends itself to a diversity of circuits rather than to the use of unit construction, and introduces most of the problems which might be encountered in a complete computer.

The paper also includes an investigation into magnetostrictive delay lines (Section 15) since these were used as storage elements.

(2) A BISTABLE CIRCUIT

The type of circuit found most suitable for performing logical operations on pulse trains is the earthed emitter connection shown in Fig. 2, since it can produce square output pulses.[1] The

Dr. Chaplin is in the Electronics Division, Atomic Energy Research Establishment, Harwell. Mr. Hayes and Mr. Owens are in the Electrical Engineering Laboratories, University of Manchester.

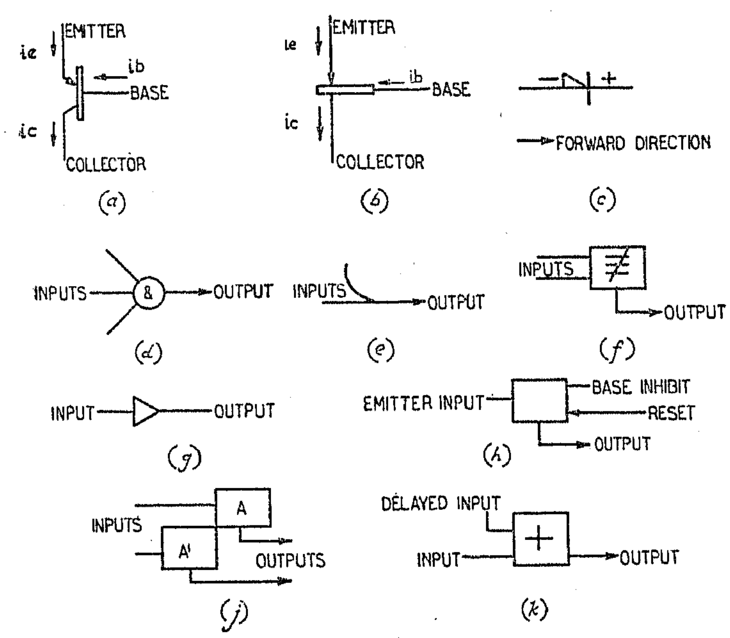


Fig. 1.—Conventions used.

(a) Point-contact transistor (LS737 or OC51).
(b) Junction transistor (OC71).
(c) Point-contact diode (CV425).
(d) "And" gate.
(e) "Or" gate.
(f) "Not equivalent to."
(g) Single digit delay.
(h) Single transistor circuit.
(j) Halver.
(k) Adder.

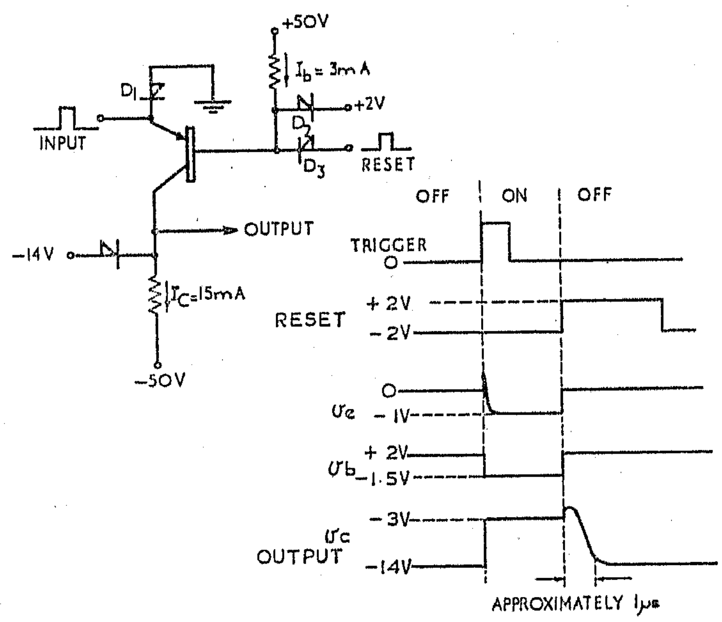


Fig. 2.—Earthed emitter bistable circuit showing "off" and "on" states and the effect of hole storage on v_c.

circuit has two states, which are stable if the following conditions are fulfilled:

In the "off" state

$$I_b > i_{co} (i_{co} \leqslant 1\,\mathrm{mA})$$

In the "on" state

$$i_c > \frac{\alpha_k}{\alpha_k - 1} I_b$$

where α_k is defined as the ratio of collector current to emitter current at the knee of the i_c/V_c curves, and varies between 2 and 4. The use of a current supply, I_b, to the base increases the value of i_{co} which the circuit will accept in the "off" state while still maintaining an emitter-to-base bias of 2 volts, and also increases the range of collector current over which the "on" state is stable. A similar current supply, I_c, is used in the collector circuit so that it can supply current in either direction to an external load without appreciably affecting the potential corresponding to the "off" state. The purpose of D_1 is to enable the circuit to be triggered into the "on" state by a positive pulse applied to the emitter, and then to present a low impedance for the subsequent emitter current.

Resetting to the "off" state is accomplished by a positive pulse applied to the base via D_3. The "hole storage effect"[2,3] makes it necessary for this pulse to be able to supply current up to the value of $(i_c - I_b)$, which in Fig. 2 is 12 mA, for up to 4 microsec with transistors of the type at present available. The problem of supplying sufficient current can be eased, wherever possible, by resetting all transistors at regular intervals by a low-impedance waveform, termed the clock. The time wasted in resetting can be reduced by the use of coils in the base circuits.

Both these features are combined in the circuit shown in Fig. 4, which is the basic trigger stage used throughout the multiplier. Before describing this circuit it is necessary to introduce the basic timing waveforms which control its operation.

(2.1) Basic Timing Waveforms

A number representing a row of 1's consists of 11-volt positive pulses of 3·5 microsec duration separated by intervals of 4·5 microsec. The regular waveform timing the numbers is called the clock (Fig. 3) and is negative for the 3·5 microsec

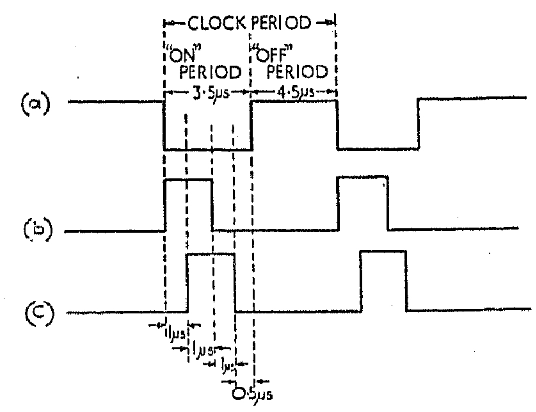


Fig. 3.—Basic timing waveforms.

(a) Clock.
(b) Strobe.
(c) Delayed strobe.
Inverse phases are also available.

digit period allowing transistors to switch on, and positive for the 4·5 microsec "off" period switching transistors off. Consecutive digits thus occur at 8 microsec intervals (the clock period) representing a pulse repetition frequency of 125 kc/s. Two other basic timing waveforms are the strobe and the delayed strobe; their timing relative to the clock is shown in Fig. 3. These three waveforms are also used in their reverse phase, in which case they are prefixed by the word "inverse." In addition to timing, these waveforms are also required to switch transistors on and off and to recharge coupling capacitors, thus relieving the transistors of merely routine loads. It is convenient to use valves for the generation of the waveforms, since only one valve for each waveform is necessary for the whole of the multiplier.

(3) THE BASIC CIRCUIT

Since the coil in Fig. 4 has negligible resistance the base potential in the "off" state is practically independent of i_{co}.

The impedance which the base circuit presents to a sharp trigger pulse at the emitter is decided largely by the damping resistor R_1, which is sufficiently high to give increased triggering sensitivity over the circuit shown in Fig. 2. Initially, the margin of stability in the "on" state is large, since i_b is not much greater than i_{co}. The voltage across the base coil (about 3 volts), how-

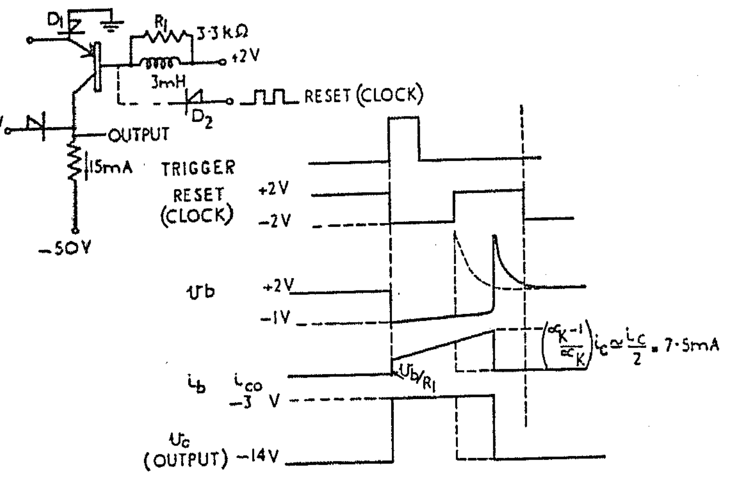


Fig. 4.—Basic trigger circuit, showing use of a base coil for self-resetting (full lines) combined with the clock for more accurate timing (broken lines).

ever, causes i_b to increase at a fairly constant rate, with a consequent gradual reduction in the margin of stability, until, when $i_b = (\alpha_k - 1)i_c/\alpha_k \simeq 7{\cdot}5$mA, the transistor reverts to the "off" state.

The duration of the positive pulse obtained at the collector depends somewhat on both i_c and the transistor parameters, but more accurate timing can be obtained by using the clock to reset the transistor at some intermediate time. This is illustrated by the broken lines in Fig. 4. Although hole storage will not have been completely eliminated at this time, it will be much reduced, with a consequent reduction of the load on the reset pulse. The duration of the reset pulse can now be reduced, since not only is the degree of hole storage reduced but the tendency of the coil to keep the high base-current flowing, as evidenced by the positive overshoot of v_b, augments the reset pulse. Resistor R_1 should be chosen to give critical damping, since the circuit will oscillate if the coil is underdamped.

The most valuable property of this basic circuit is that, when triggered, it emits an output pulse of well-defined amplitude and duration, which is sensibly independent of the load into which it operates and of the nature of the trigger pulse. In fact, the output impedance in the "on" state is largely determined by the forward impedance of D_1. Before describing how this circuit can be used for gating or delaying pulses, the following Sections discuss some alternative resetting methods.

(3.1) Alternative Resetting

It is essential, in Fig. 4, that the reset pulse shall not exceed +2 volts, unless it is of shorter duration than the positive overshoot of v_b; otherwise, if it exceeds +2 volts it may establish a reverse current in the base coil, and the transistor may trigger "on" when the reset pulse is removed. Although the top level of the clock pulses, derived from a valve, can be controlled within the required limits, a reset pulse originating from the collector of another transistor may be uncertain in its top level to the extent of 2 volts. In this case a secondary winding of, say, three times as many turns can be added to the base coil, as shown in Fig. 5(*a*). The uncertainty in the base potential during reset is thereby reduced by two-thirds, as is also the load on the reset pulse. Alternatively, the reset pulse may be applied to the fixed end of the base coil to increase the voltage across it (and hence the rate of rise of base current) by a factor of 3 or 4 [Fig. 5(*b*)]. This arrangement suffers from the resulting delay in switching off and the necessity for generating the reset pulse at high level, but it has an application discussed in Section 8.2. A similar effect could be obtained by the use of saturable magnetic material in the base coil.

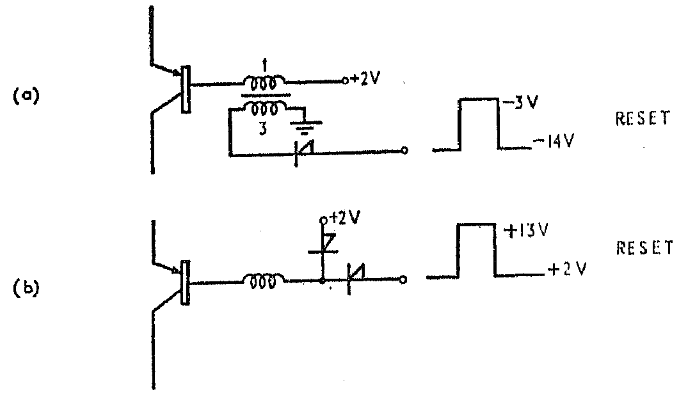


Fig. 5.—Alternative resetting methods.
(*a*) Transformer method.
(*b*) Series method.

(3.2) Gating

The problem of gating is basically that of arranging that a controlling waveform shall allow or inhibit the passage of a pulse train. In the valve circuits this controlling waveform is often derived from a cross-coupled 2-state circuit or staticizor, which has two output points of opposite phase [Fig. 6(*a*)]. For a given

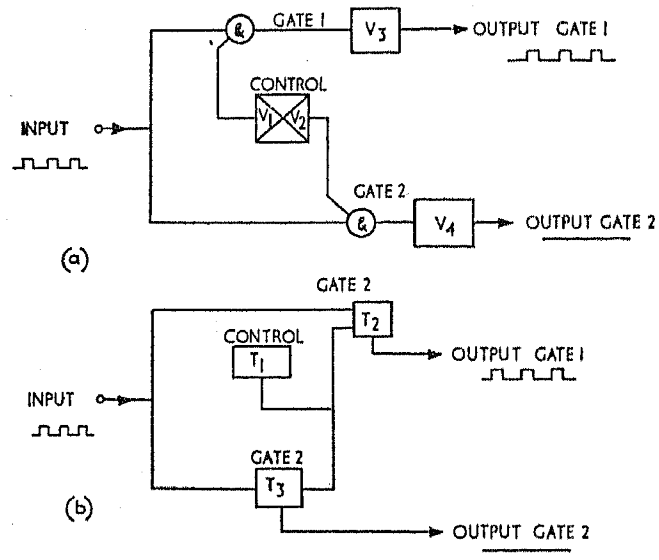


Fig. 6.—Direct gating.
(*a*) With valves.
(*b*) With transistors.

state of this staticizor the pulse train can be allowed or prevented by connecting to one or the other of the output points. At first sight, the fact that a single-transistor 2-state device has only one output point might appear to be a drawback, but fortunately, although it has only one output point, the transistor trigger circuit has two input points which are in antiphase with respect to the effect they have on the output. Thus the pulse train is allowed or prevented, depending on which of the two input points to the gate the control waveform is connected to [Fig. 6(*b*)]. These two input points are the emitter and the base, and both of them have current and voltage gain to the collector, so that one such circuit can control several others. The complete gating circuit is shown in Fig. 7.

The emitter gating circuit is enclosed by the left-hand dotted line. An input pulse to the circuit, from another such collector, will cause R_1 to pass about 3 mA, which will reach the emitter and trigger the circuit only if there is a coincident positive pulse on the cathode of D_1 (the control waveform). To prevent triggering, the control waveform has to absorb this current, and so, if its collector supply current amounts to 15 mA, it can control up to 4 similar gates.

The base-gating circuit is enclosed by the right-hand dotted line. A positive pulse (the control waveform) on the cathode of D_4

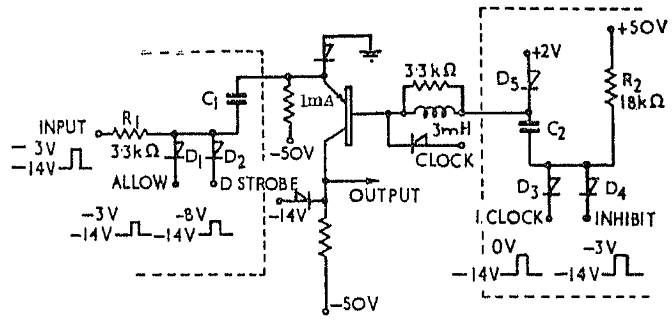


Fig. 7.—Basic circuit with gating facilities.
Left: Positive control "allows."
Right: Positive control "inhibits."

allows the current flowing in R_2, which is greater than i_{co}, to raise the potential of the whole of the base circuit through about 11 volts to +13 volts. Since D_2 prevents the input from raising the emitter above +6 volts the base is at a higher potential than the emitter and the circuit will not trigger. If the current flowing through R_2 is 3 mA, then, as in the previous case, the controlling waveform will be able to control up to 4 such circuits.

Thus a positive pulse on the cathode of D_1 allows triggering, whereas a similar pulse on the cathode of D_4 prevents it. Coupling capacitors C_1 and C_2 can be regarded as 14-volt and 16-volt batteries respectively, since they are large enough to maintain their potential difference over a digit period, and any loss of charge is made up by the delayed-strobe and inverse-clock waveforms respectively in the "off" periods. Reference to Fig. 3 shows that these waveforms also ensure that the base inhibit always overlaps any waveform appearing at the emitter, and so prevents spurious triggering.

(3.3) Delay

It is often required to retime a pulse to the beginning of the next digit period, regardless of the particular part of the previous digit period in which it occurred. The basic trigger circuit can be adapted to perform this function as shown in Fig. 8. A pulse

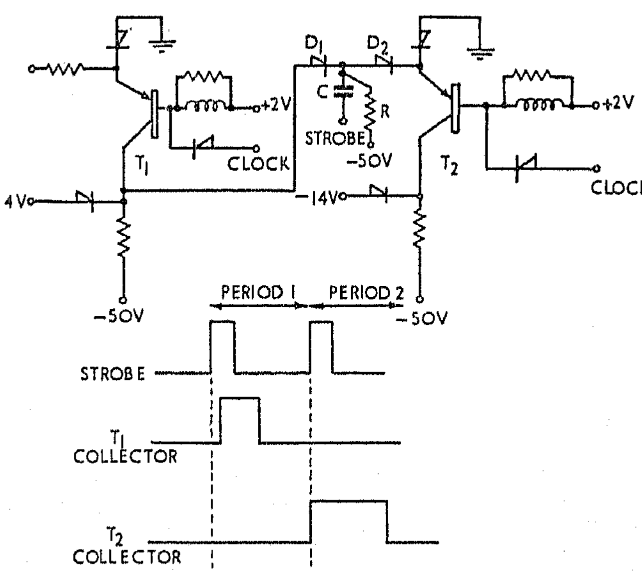


Fig. 8.—Basic circuit with retiming facilities.

appearing at the collector of T_1 some time in the first digit period charges the small capacitor C, which is then discharged by the strobe into the emitter of T_2 at the beginning of the next digit period.[1]

The basic trigger circuit shown in Fig. 4 can thus have attached to it either or both of the gating systems of Fig. 7 or the delay system of Fig. 8, and if suitable precautions are taken it can embody both gating and delay simultaneously. There is no interaction between these systems.

(3.4) Indirect Gating

So far, only instantaneous gating has been discussed, but if sufficient time is available between each control and input pulse both allowing and inhibiting can be achieved without using the base.

An input pulse as in Fig. 9(*a*) can cause D_1 to conduct, and hence to trigger the transistor, only when the control waveform

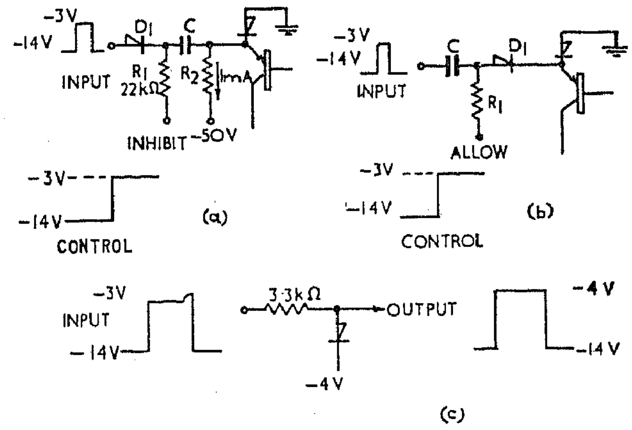


Fig. 9.—Indirect gating.
(*a*) Positive control "inhibits."
(*b*) Positive control "allows."
(*c*) Input clipping circuit.

is at −14 volts. Thus a positive control waveform inhibits the circuit. On the other hand, the input pulse in Fig. 9(*b*) can trigger the transistor only when the control waveform is positive. Thus a positive waveform allows an input to trigger the transistor.

In either case the time required to set the gate is an appreciable multiple of CR_1. The minimum capacitance of C for reliable triggering is about 500 $\mu\mu$F. If the current flowing through R_1 in Fig. 9(*a*), when the gate is being set to inhibit, exceeds that in R_2 the transistor may be triggered by the control waveform. To prevent this, R_1 should not be less than 15 kilohms. Thus the minimum value of CR_1 is 7·5 microsec.

To prevent spurious triggering due to any discrepancy in the top levels of the input and control waveforms, a diode clipping circuit [Fig. 9(*c*)] is placed in series with the inputs of the gating circuits in Figs. 9(*a*) and 9(*b*).

(3.5) Binary Counter

The waveforms which control the operation of the multiplier are generated by a series of binary counters. Single-transistor binary counters of the earthed-emitter type can be constructed, in which the input is gated to both emitter and base by the indirect gates of Figs. 9(*a*) and 9(*b*). This, however, involves resetting the transistor through a capacitor, which must be relatively large and therefore costly in both time and power. The two-transistor circuit of Fig. 10 is used in preference.

The lower transistor, T_2, which supplies the output v_{c2}, is switched on by the rising edge of the input (via the lower indirect gate) only if its collector has been negative. Similarly, T_1 is switched on via the upper indirect gate only if the collector of T_2 has been positive, and its function is to reset T_2. Since direct coupling is used for resetting T_2, all the bias potentials of T_1 are raised by 6 volts. The base coil of T_1 need only have inductance high enough to allow T_2 to be reset, whereas the base coil of T_2 must be large enough for T_2 not to be self-resetting. Two outputs are available from this circuit: a symmetrical waveform, v_{c2}, and short positive pulses at high level, v_{c1}, which can be used for resetting other circuits. The circuit can be operated by either v_{c1} or v_{c2} of the previous binary counter, but if by v_{c1} the resistance R_1 must be added and the cathode potential of D_1 increased to +2 volts.

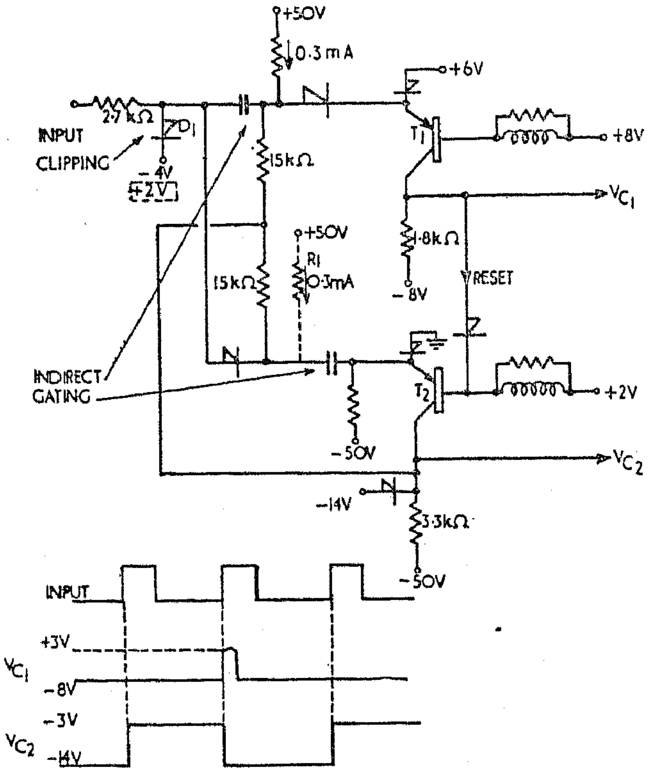


Fig. 10.—Two-transistor halving circuit.

(4) MULTIPLIER LOGICAL SYSTEM

Before the method of using these basic circuits in the multiplier is described, it is necessary to outline its logical system.

The two numbers, the multiplicand x and the multiplier y, which are to be multiplied together, are each 32 digits long. In order that the multiplier should provide a comprehensive test for the transistors with a minimum outlay of equipment, it is arranged that the number y is split up into four parts of eight digits each, y_A, y_B, y_C and y_D, and x is multiplied by each part in turn.[6,7] This reduces the required number of adders from 31 to 8. In addition to this, the digits of y are taken in pairs,[4] resulting in a further reduction in the number of adders required from 8 to 5. This also has the advantage, in this case, of requiring a more complex gating system for selecting the numbers which are fed into the adders.

An example of paired multiplication in which x and y are both 8 digits long is shown in Fig. 11. The numbers are written with

x	1 0	1 1	0 1	1 1			
y	1 1	0 0	1 0	0 1			
$3x \times 2^0$	1 1	1 0	0 0	1 1	0 1		
0×2^2		0 0	0 0	0 0	0 0		
$x \times 2^4$			1 0	1 1	0 1	1 1	
$2x \times 2^6$				0 1	0 1	1 0	1 1 1
Product xy	1 1	1 0	1 0	0 0	0 0	0 1	0 0 0 1
$0 =$	0 0	0 0	0 0	0 0			
$x =$	1 0	1 1	0 1	1 1			
$2x =$	0 1	0 1	1 0	1 1 1			
$3x = x + 2x =$	1 1	1 0	0 0	1 1 0 1			

Fig. 11.—Example of paired multiplication.

the least significant digit on the left, as they would appear on a cathode-ray tube.

Considering the least significant pair of digits of y, since there are four possible combinations of these two digits, there are four possible results of multiplying x by them. These results are 0, x, $2x$ and $3x$, and the appropriate one in this case is $3x$, since both digits are 1's. The next most significant pair of digits of y are both 0's, so their contribution is zero, but displaced two digits to the right. The next pair is 1 0 and its result is x displaced a further two digits to the right. The final pair, 0 1, has the result $2x$ displaced by another two digits. These numbers, x, $2x$ and $3x$ are generated separately and the appropriate one is gated into the adders by each pair of digits of y in turn. This is accomplished by the arithmetic unit, which is described in detail in Section 8.

The arithmetic unit represents about one-third of the total equipment involved. The other two-thirds is divided fairly equally, first between the generation of numbers and timing waveforms, and secondly the necessary stores with means for selecting the appropriate numbers.

A simplified block diagram of the whole multiplier is shown in Fig. 12. The pulse separator is operated from the basic clock

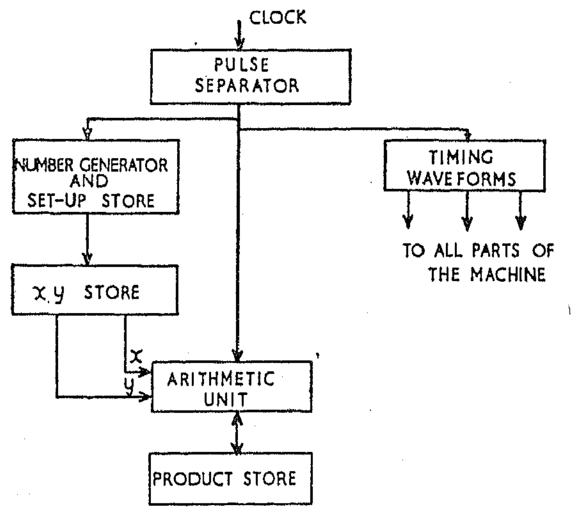


Fig. 12.—Basic block diagram of multiplier.

waveforms, and provides means for identifying each particular digit period, and hence any digit of the number. The numbers x and y are generated from the pulse separator and then transferred via a set-up store to the xy store, and the system is so arranged that multiplication begins automatically with the transfer of the second number, y, to the xy store. When this happens, x and the appropriate portions of y are gated into the arithmetic unit, and when the multiplication is complete the circuits are automatically reset to their initial states. The product, xy, is stored in the product store. Timing waveforms are derived from the pulse separator mainly by the process of counting down, and are applied to all parts of the multiplier.

It is proposed to consider each part of the multiplier in turn, to explain its logical operation in more detail, and to describe any transistor circuits that are markedly different from the basic circuits described in Section 3.

(5) PULSE SEPARATOR

The first unit of Fig. 12 to be described is the pulse separator, which is shown in Fig. 13 and consists of eight of the delay stages

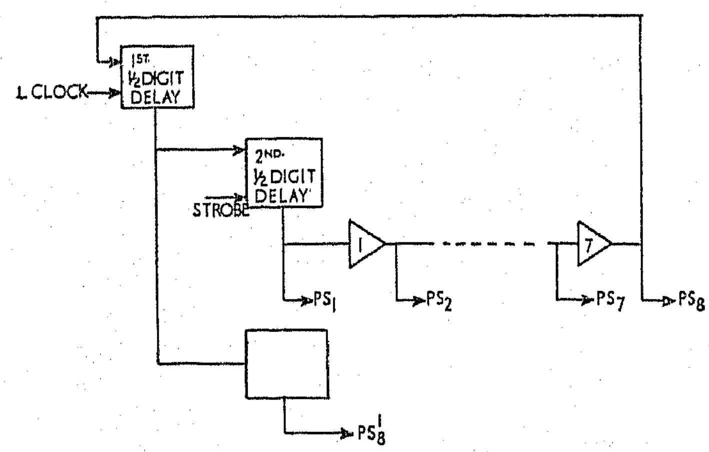


Fig. 13.—Block diagram of pulse separator.

of Fig. 8 connected in a ring. The output from any one stage is thus one pulse in every eight digit periods, and, if the stages are labelled 1 to 8, any one stage defines a particular digit period in the complete cycle of eight periods. Stage No. 8 actually consists of two half-digit delay stages, the junction of which defines the beginning of the 4·5 microsec "off" period immediately preceding the No. 1 digit period. This is used to generate PS_8', a high-level self-resetting transistor waveform, which initiates all the main timing waveforms. This ensures that they are well established before the first digit period.

The two half-digit delays are achieved by charging a capacitor and then discharging it with the appropriate timing waveform, and they also incorporate the means of ensuring that one (and only one) pulse circulates.[10]

(6) TIMING WAVEFORM GENERATOR

Timing waveforms having periods ranging from 8 to 416 digit periods are necessary to control operations (Fig. 14). Wave-

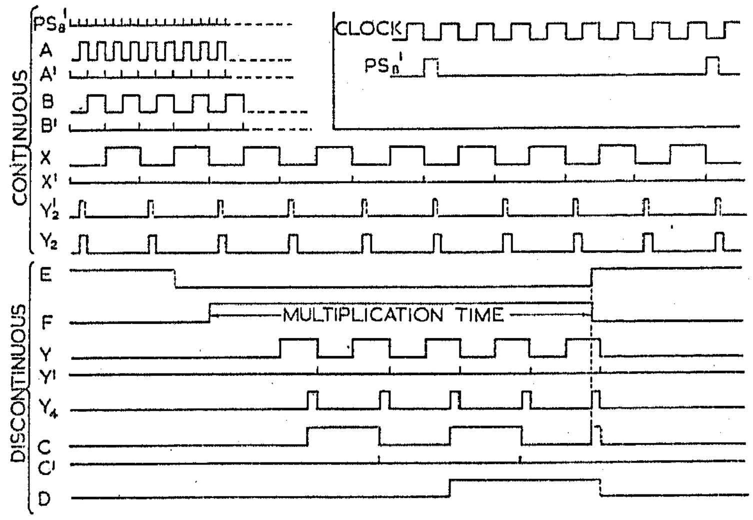


Fig. 14.—Timing waveform diagram.

forms PS_8', A, B, X and Y_2 run continuously and are timed to the clock waveform which is shown inset with PS_8' on an expanded scale. Their main purpose is to allow numbers to be set up prior to multiplication. Some of these, such as A and B, are symmetrical and are obtained by halving, using the circuit shown in Fig. 10. Thus A and A′ are halved from PS_8' and B and B′ are halved from A′. Other waveforms, such as Y_2, are unsymmetrical and are generated from combinations of previously derived waveforms by direct and indirect gating on the emitter and resetting on the base. This is illustrated by the circuit shown in Fig. 15, in which Y_2 and Y_2' are generated. Y_2 is generated by T_1 by passing X and B through the "or" gate of D_3, D_4 and R_3 and using the resulting "X or B" to control the indirect gate of R_2, D_2 and C_1 which has A as its input. The input to the emitter is thus "A not (X or B)," and the output from the collector is a pulse which is triggered by the input but reset by the base coil. The duration of this pulse, Y_2', will depend on the load into which it operates, and this fact is utilized in the method of resetting the staticizors described in Section 8.2. The fact that the indirect emitter gate does not respond immediately is an advantage in this case, since it prevents spurious inputs being applied to the emitter of T_1 on account of the transitions of A, X and B not being coincident.

The waveform Y_2 is generated from Y_2' by T_2, which is the basic circuit of Fig. 4 triggered by Y_2' and reset by A′, a high-level waveform.

The rest of the waveforms in Fig. 14 operate only during multiplication, and their function is to initiate multiplication, to select the appropriate parts of y in the correct order, and to

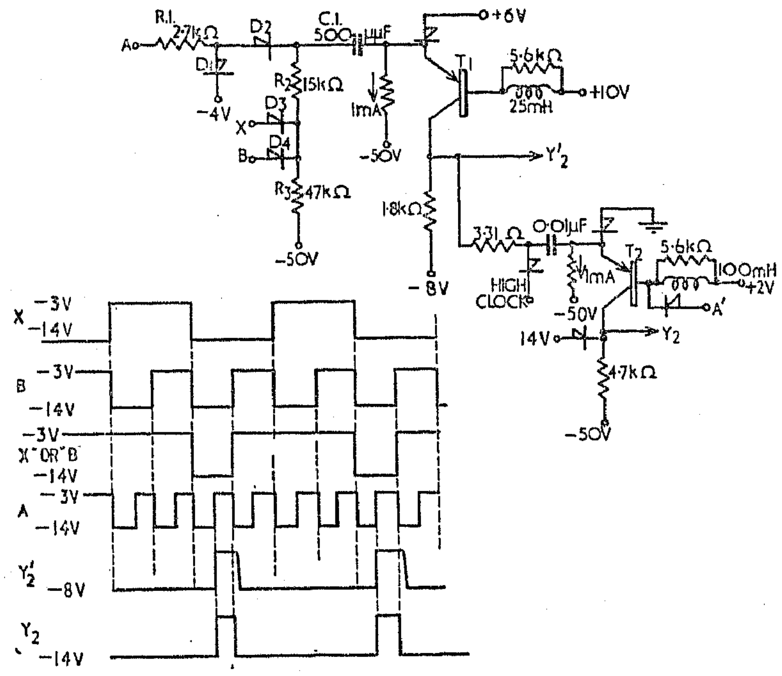


Fig. 15.—Generation of waveforms Y_2' and Y_2.

terminate multiplication and allow the product to remain circulating in the product store. Multiplication is initiated by a single pulse occurring at a random time, which switches off the transistor that generates E, as shown by the broken line. Since E may be on for an indefinite period it has a resistive base load. When E falls it removes the inhibit on F, which is then switched on by the next X′ pulse. Thus F rises at a standard time and controls the multiplication in the following manner.

The waveform Y is triggered by "X′ and F" and reset by Y′, which in turn is triggered by "B′ and Y" and is self-resetting. Y_4 is triggered by "A and B and Y" and reset by A′. Waveforms C, C′ are generated by the standard halver triggered by Y_4, and D is triggered by the second positive-going edge of C. The third positive-going edge of C allows the next PS_2 pulse to trigger E, which in turn resets C, F and D by the delay-line method described in Section 8.2, causing Y, Y′ and Y_4 to terminate, thus completing the multiplication. The end of multiplication is timed by PS_2 instead of PS_1 to allow for the extra delay in the x, $2x$ and $3x$ generator (Section 8.4).

(7) NUMBER GENERATOR AND SET-UP STORE

The number generator and the set-up store are not strictly part of the multiplier, but simulate the external computer of which the multiplier might be a part.

The procedure is first to set up the number x in the 32-digit-long set-up store and manually to gate it into the 64-digit-long xy store using waveform X. The set-up store is then cleared and the second number, y, formed. On starting multiplication, y is automatically transferred to the vacant place in the xy store by a transistor gate which is controlled by "F and Y" on the emitter and inhibited by "X or Y" on the base. Reference to the waveform diagram of Fig. 14 shows that this gate is open for only one Y period, resulting in a "once only" transfer.

The method of generating the numbers is shown in Fig. 16. A series of switches, S_1 to S_8, connects the collectors of the eight pulse separators, via diodes, to the junction of R_1 and R_2, where an 8-digit number can be set up. This number is fed into the set-up store via two 2-way switches, the four possible configurations of which decide the particular quarter of the Y period into which the number is placed. The first switch, switch a, selects one of two clamping waveforms, Y_2 or B, which, in conjunction with the clamping waveform A, define the periods Y_2 or Y_4, respectively. Reference to the timing waveform diagram of

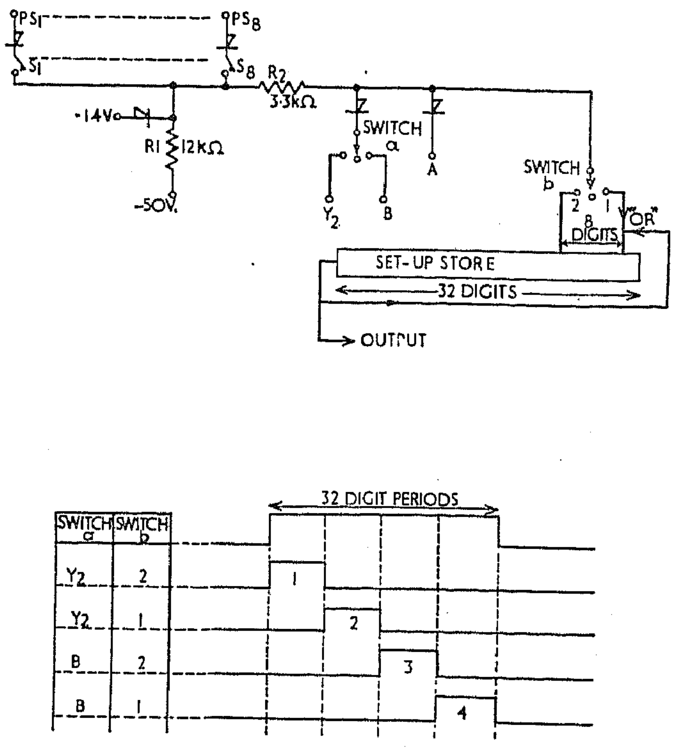


Fig. 16.—Number generator with set-up store.

Fig. 14 shows this to be so, since "A and B" is a waveform similar to Y_4, but of twice the frequency, and "A and Y_2" is identical to Y_2. Y_4 could not be used directly, since it operates only during multiplication.

If switch *b* is in position 1, then the periods Y_2 and Y_4 are defined, but if it is in position 2 the numbers are inserted eight digit periods in advance into the store, and periods Y_1 and Y_3 are defined.

It is important that the lead to the left-hand side of R_2 be carefully situated, since a temporary short-circuit to, say, -14 volts would burn out all eight pulse separators.

Details of the set-up store are given in Section 9.5.

(8) ARITHMETIC UNIT

(8.1) Logical Operation

The actual multiplication is accomplished in the arithmetic unit, a block diagram of which is shown in Fig. 17.

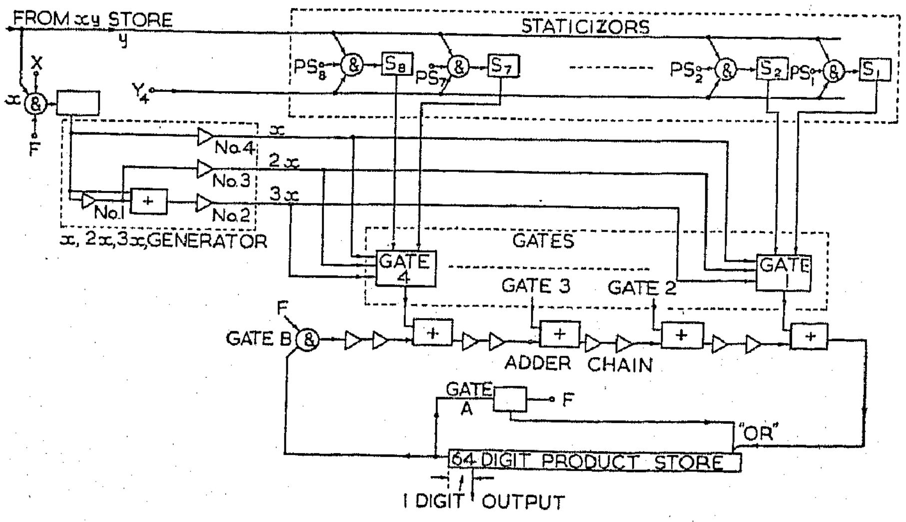


Fig. 17.—Block diagram of arithmetic unit.

The eight staticizors at the top of the diagram are set up to conform to the first batch of eight digits of *y*, using the pulse separator and Y_4 for timing. The result is that y_D is set up, since it occurs in period Y_4. Each gate controlled by its pair of staticizors then allows 0, x, $2x$ or $3x$ to flow into its adder during the following x period. The resulting partial product $(x \times y_D)$ then enters the product store. In the next Y_2 period all the staticizors are reset by Y_2' and are then set up to the next batch of y, (y_C), in period Y_4 and the next partial product $(x \times y_c)$ is obtained and added to the first one, which will just have emerged from the product store. This is repeated for y_B and y_A and the resulting final product is kept circulating in the product store by opening gate A and closing gate B.

Before multiplication, y_A, y_B, y_C and y_D occur at the times Y_1, Y_2, Y_3 and Y_4 respectively, but during multiplication y is progressively shifted 8 digit periods to the right, as shown in Fig. 18. The reason for this is that to set up the staticizors to

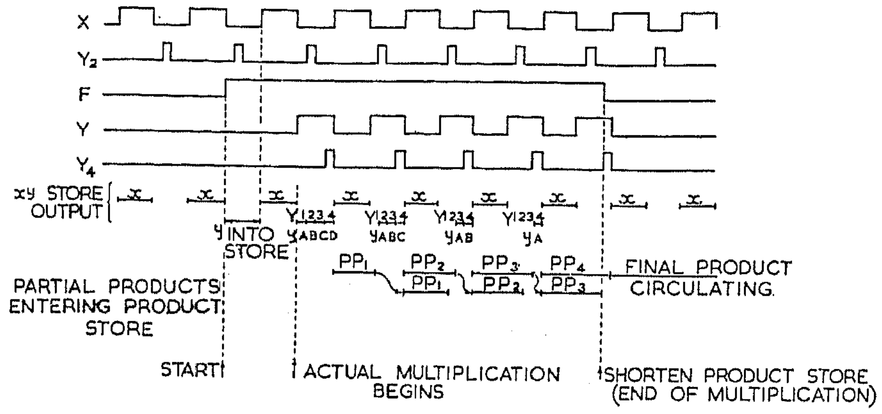


Fig. 18.—Multiplier timing and sequence diagram.

each quarter of y in turn it is convenient to look at the same place (the xy store output) at the same time (Y_4) and see the four quarters of y appearing in turn. This avoids either switching to four different points in the store or generating an irregular timing waveform.

The choice of which part of y to look at is decided as follows. Period Y_1 would not allow sufficient time to reset the staticizors, and the two middle periods Y_2 and Y_3 would involve taking $y_{A,B,C,D}$ in an irregular order.

(8.2) Circuit Details

(8.2.1) The Staticizor.

Each staticizor is the single transistor circuit of Fig. 2 switched on by "y and Y_4 and PS_1, etc.," and reset by Y_2'. Base coils are not used because the ratio of time-on to time-off may be as high as 3 to 1, which would not allow sufficient time for the coils to recover.

If Y_2' were to reset all eight staticizors simultaneously it would have to supply a current up to $8 \times (I_c - I_b) = 96\,\mathrm{mA}$, whereas it is required to limit this current to 30 mA for reliable operation. Since plenty of time is available, this can be achieved by the circuit of Fig. 19, which shows the Y_2' transistor of Fig. 15 connected directly to the base of the first staticizor and via a series of coils to the others. When Y_2' rises, the first base is taken to $+2$ volts and draws a current of $(I_c - I_b) = 12\,\mathrm{mA}$. This current gradually decays with the hole storage, but current begins simultaneously to flow through L_1, since its left-hand side is now at $+2$ volts and its right-hand side is at about -2 volts (assuming all the staticizors are on). This current increases almost linearly until, at 12 mA, the potential of the second base rises to $+2$ volts. This action then proceeds along the chain of staticizors until all are off. During this time the current from Y_2' has remained substantially constant at $i_{D1} + i_{D2} \ldots i_{D8}$ ($\simeq$ 12 mA), but it now falls to zero and the ensuing reduction in the collector current of Y_2', together with the action of its base coil, causes it to reset.

The operation of the circuit is analogous to a lumped delay line (ladder network) in which the capacitors are replaced by current-operated 2-state devices, the reset pulse being propagated down the line.

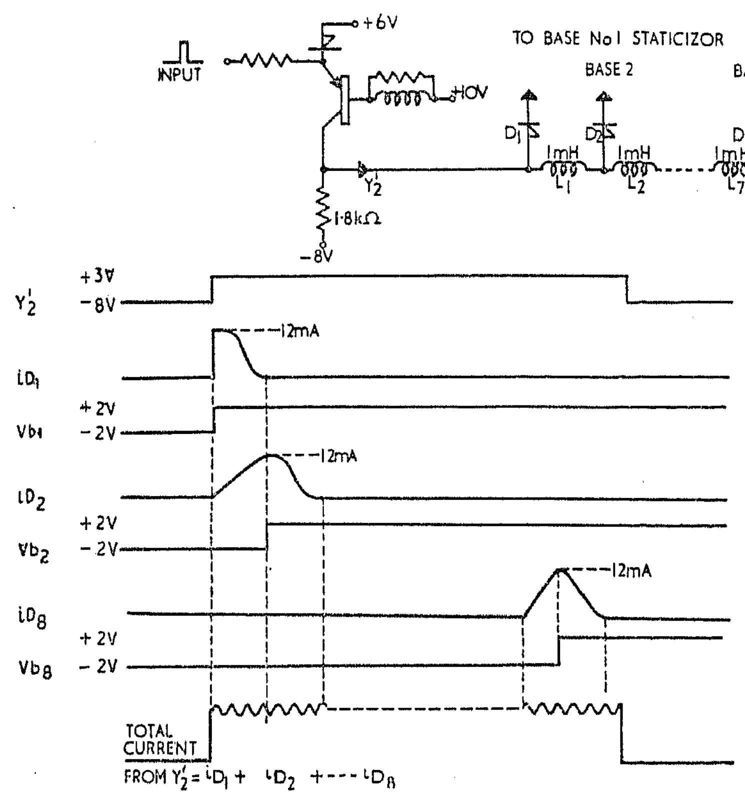


Fig. 19.—Circuit and waveforms for resetting staticizors.

(8.2.2) Three-Input Gates.

The function of each gate in Fig. 17 is to allow one of the four numbers 0, x, $2x$ or $3x$ through into the adder chain, according to the configuration of its two controlling staticizors. The circuit diagram of the first gate, which is controlled by the first two staticizors, is given in Fig. 20.

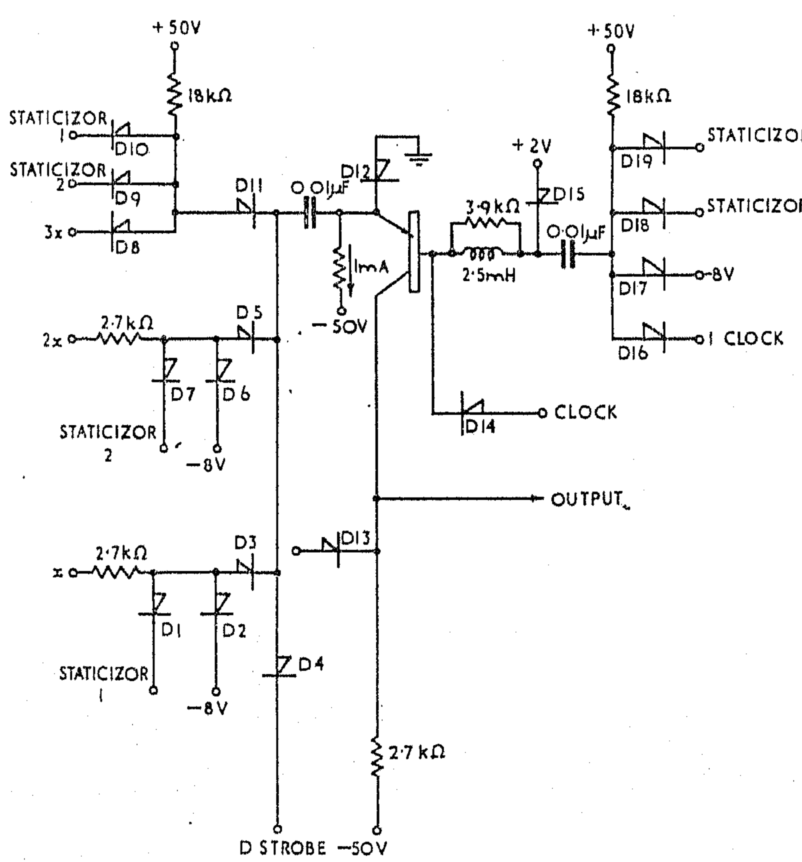


Fig. 20.—Three-input gate with two control waveforms.

Basically, the circuit is the standard gate, having provision for inhibiting on both emitter and base. The x and $2x$ inputs are applied to the emitter via current-limiting resistors and diodes, and the $3x$ input via a diode gate, the method of selection being as follows:

(i) *Both Staticizors off.*

The cathodes of D_1, D_7, D_9, D_{10}, D_{18} and D_{19} are clamped at -14 volts and so no input can reach the emitter. There is therefore no output.

(ii) *Staticizor* 1 *on, Staticizor* 2 *off.*

The cathodes of D_7, D_9 and D_{18} are at -14 volts, but the cathode of D_1 is raised to about -3 volts. The number x is thus able to raise the anode of D_1 from -14 to -8 volts, where it is caught by D_2, producing a 6-volt input at the emitter, sufficient to overcome the 2-volt bias on the base. The output is therefore x.

(iii) *Staticizor* 2 *on, Staticizor* 1 *off.*

In this case the cathodes of D_1 and D_{10} are clamped and that of D_7 raised, allowing $2x$ to produce a 6-volt waveform at the emitter. The output is therefore $2x$.

(iv) *Both Staticizors on.*

The cathode potentials of all the diodes, D_1, D_7, D_9, D_{10}, D_{18} and D_{19} are raised, allowing both x and $2x$ to raise the emitter to $+6$ volts. At the same time, however, the fact that the cathode potentials of D_{18} and D_{19} are raised enables the base to rise through 6 volts until effectively caught by D_{17} at $+8$ volts, preventing either x or $2x$ from producing an output. The number $3x$, however, is free to raise the emitter to about $+11$ volts, which exceeds the base potential and produces an output of $3x$. The $3x$ input is provided with a diode and current source instead of simply a limiting resistor because of the large voltage excursion of the emitter. The inverse clock and delayed strobe are, as usual, applied to base and emitter, respectively, to recharge the coupling capacitors at each digit period.

(8.3) The Adder Chain

The adder chain consists of alternate adders and double delays. Since each delay stage is the single-transistor circuit of Fig. 8, the double delay uses two transistors, and it might appear advantageous to replace them by a two-digit-long electrical delay line. One of them, however, is necessary, since it must drive the following adder, and a delay line could not do this. Although

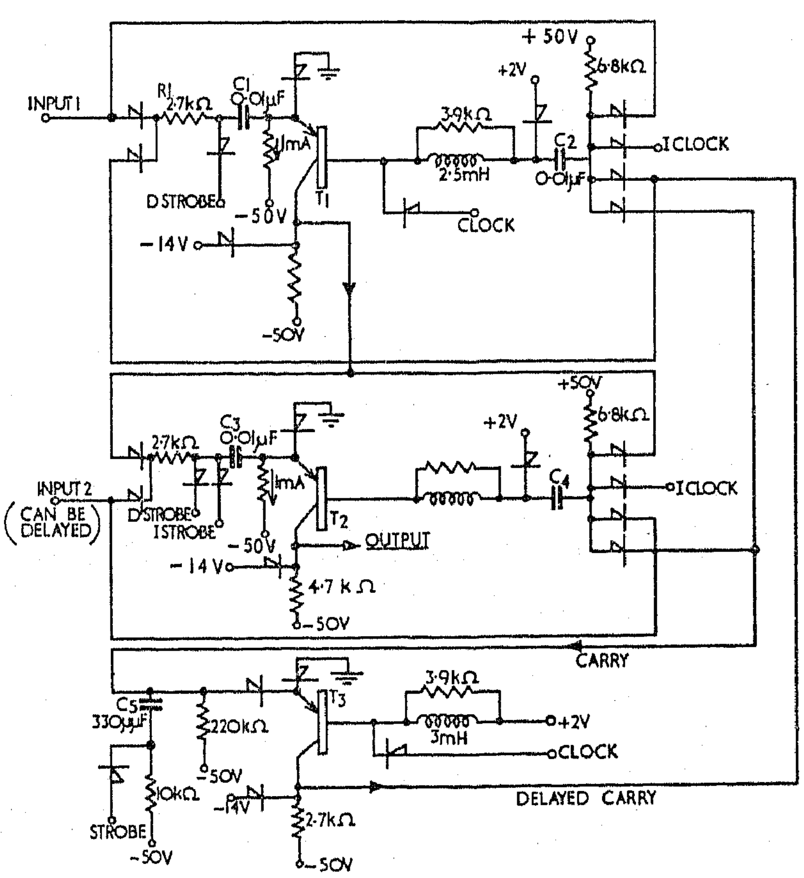


Fig. 21.—Adder circuit.

the other one could be replaced by a single-digit-long line it has been retained because it is less bulky.

Each adder is of the logical type and is shown in Fig. 21. The first two stages, T_1 and T_2, are based on Fig. 7 and are identical. Each has two inputs, and produces an output pulse only when the two input pulses are different, i.e. when one is 0 and the other 1. Since the output from T_1 is one of the inputs to T_2, the T_2 stage must incorporate some delay to ensure that both inputs are present before an output is attempted. This is accomplished by the combination of the delayed-strobe and inverse-strobe waveforms in the emitter circuit of T_2. The final stage, T_3, delays the "carry" by one digit period and is similar to one of the stages of Fig. 8.

Since the operation of T_2 is delayed with respect to that of T_1 it can accept a delayed input, and for this reason the input from the gates (Fig. 17) which occurs late in the digit period is connected to T_2, and that from the delays is connected to T_1.

(8.4) The x, $2x$ and $3x$ Generator

The method of generating x, $2x$ and $3x$ (Fig. 17) is to gate x from the xy line by waveform "X and F," to delay x by one digit period in delay No. 1, forming $2x$, and to add x and $2x$ in the adder to form $3x$. The x gate employs a transistor, since it must drive the adder. The output from the adder occurs late in the digit period and so must be retimed by delay No. 2 before it can operate the gates. To compensate for this extra delay on $3x$, delays Nos. 3 and 4 are inserted in $2x$ and x, respectively, with the additional advantage that the transistors which operate the gates (delays 2, 3 and 4) have no other load imposed on them.

(9) THE STORAGE SYSTEM

Magnetostrictive delay lines[8,9] were chosen as storage elements, because they offer many advantages over other forms at the operating frequency of the multiplier, stores using these delay lines being physically robust and cheap and easy to construct. The delay elements themselves have the advantage of being insensitive to external electric and magnetic fields and to normal changes in temperature. They provide an easily adjustable delay, and access is possible at intermediate points.

The delay-line element is basically as shown in Fig. 22. It

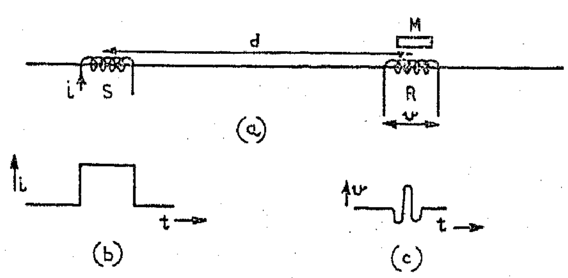


Fig. 22.—Delay line.

consists of a length of nickel wire with two coils, separated by a distance d, wound upon it. A pulse of current, as shown in Fig. 22(*b*), is fed into the sending coil S, which causes a contraction of the material immediately inside the coil. The resulting disturbance is propagated along the wire as a stress wave with a velocity $c = E/\rho$, where E is Young's modulus, ρ is the density and c is the velocity of sound in nickel, and arrives at the receiving coil R at a time t ($= d/c$) later. The permeability of the nickel at R is changed by stress, and so the stress wave causes the induction threading the coil R (due to the magnet M) to change momentarily, resulting in an output pulse, as shown in Fig. 22(*c*).

Fig. 23 is a block diagram of the system used to regenerate and retime the output pulses and feed them back into the input. It comprises an amplifier for raising the level of the received

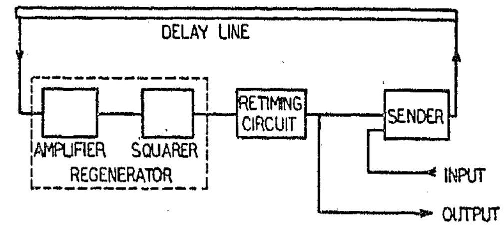


Fig. 23.—Block diagram of regenerative loop.

pulses sufficiently to actuate a trigger stage (the retimer), a squaring stage to isolate the amplifier from the retimer, a retiming stage to synchronize the circulating pulses to the clock, and a sending stage to convert the output of the retimer into current pulses for injecting into the line.

(9.1) The Sending Circuit

The sending circuit (Fig. 24) is the basic regenerative amplifier of Fig. 4, except that the sending coil is in series with the collector and the collector supply current has been increased.

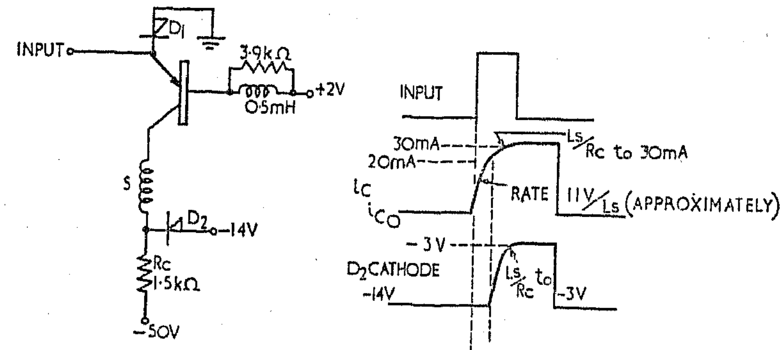


Fig. 24.—Sending circuit and waveforms.

The input pulse switches on the transistor and impresses 11 volts across the coil S. This coil is of relatively small inductance and the current flowing in it rapidly increases to 24 mA. When this happens D_2 cuts off and its cathode rises to bottoming potential as the current in the coil continues to increase to its maximum value of 33 mA. The duration of the "on" state is defined by the base coil, which resets the transistor, causing the collector current to fall rapidly to i_{c0}. Since the rise-time of the collector current is less than 1 microsec the resulting current pulse of 3 or 4 microsec is approximately square.

Gating on emitter and base, as in Fig. 7, and delay, as in Fig. 8, can be effected on the sending transistor with no adverse effect on its operation. Thus gating, retiming and the formation of a current pulse can be accomplished by a single transistor simultaneously, if required.

(9.2) The Amplifier

The receiving coil R has an internal impedance of 300 ohms and supplies a short-circuit current of about 20 μA. To operate the squaring stage this signal must be increased to 5 mA at a similar impedance. The amplifier must therefore have a current and voltage gain of 250. The output from the receiving coil is shown in Fig. 25(*a*). A single 1 consists of a full-amplitude positive pulse between two half-amplitude negative pulses, the

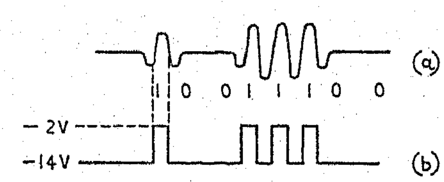


Fig. 25.—Regenerator waveforms.
(*a*) Amplifier input.
(*b*) Squarer output.

combination being a balanced waveform. Adjacent 1's thus consist of alternate positive and negative full-amplitude pulses approximating a sine wave. The frequency spectrum of the waveform extends either side of the pulse repetition frequency of 125 kc/s, from about 25 kc/s to 225 kc/s.

The choice of type of amplifier to be used (earthed-base, emitter, or collector) is easy, since the point-contact unit tends to be unstable if the base is not earthed, and junction units at present available have no appreciable gain in any but the earthed-base connection at the required frequency. The junction unit was chosen in preference to the point-contact unit because of its more uniform current gain. The amplifier consists of three earthed-base transformer-coupled stages. For low-frequency operation under matched conditions, transformer ratios of about 30/1 are required, since the ratio of R_{22}/R_{11} is approximately 900. Such an amplifier, however, would have a very poor frequency response, each stage cutting off in the region of 30 kc/s. This is caused primarily by an effective *RC* time-constant in the collector equivalent circuit, and by an effective base resistance which increases with frequency, although the short-circuit current gain, α, of the junction transistor remains above 0·9 up to at least 600 kc/s. An alternative method is to operate the transistors as quasi-current amplifiers with such small loads that the collector time-constant may be disregarded. The frequency response of the amplifier may then approximate to that for α. The power gain per stage is of course very much reduced by this method and a compromise must be found between the conflicting requirements of enhanced frequency response and a useful value of power gain.

A collector load of 5 kilohms was chosen, since it extends the gain of the complete amplifier well above 300 kc/s, and allows a reasonable value of power gain.

Under these conditions the amplifier consists of three transformers having ratios, and hence current gains, of 7/1, 7/1 and 5/1 respectively, matched to each other by transistors each having a current gain of almost unity. The current or voltage gain of the amplifier is thus defined by the three transformers as $7 \times 7 \times 5 = 245$, resulting in a power gain of 245^2, and the design of the amplifier is reduced to the design of suitable transformers.

(9.3) The Squarer and Retimer

The amplifier output must eventually operate a trigger stage, but it is desirable to present to it a sensibly linear load because of its a.c. couplings. It is also desirable to square the output from the amplifier to produce a signal of fixed amplitude, as shown in Fig. 25(*b*).

The squaring stage, T_1 in Fig. 26, performs three functions. First it presents a fairly linear load to the amplifier, since the emitter of T_1 conducts when the input is positive and conducts D_1 when it is negative, resulting in an input impedance of approximately R_1. Secondly, it produces a square output having a fixed amplitude even though the input amplitude may vary. Thirdly, spurious signals due to unavoidable reflections from the terminations of the delay line are biased off by the small positive voltage on the base. A smaller transformer ratio of 5/1 is used for the squarer to allow for its higher input impedance. Since a junction transistor is used in the squarer its collector supply current need be no greater than 100 μA to bias off i_{c0}, but it is essential to include diode D_2 to prevent bottoming, otherwise the transistor will not switch off in time.

The output from the squarer is fed directly into the standard delay stage, T_2 in Fig. 26, for retiming to the beginning of the next digit period. The delay of the line is adjusted so that signal pulses arrive back midway between two adjacent strobe pulses. There is thus a half-digit-period margin of safety in either direction to guard against variation of delay with temperature and variation of clock frequency.

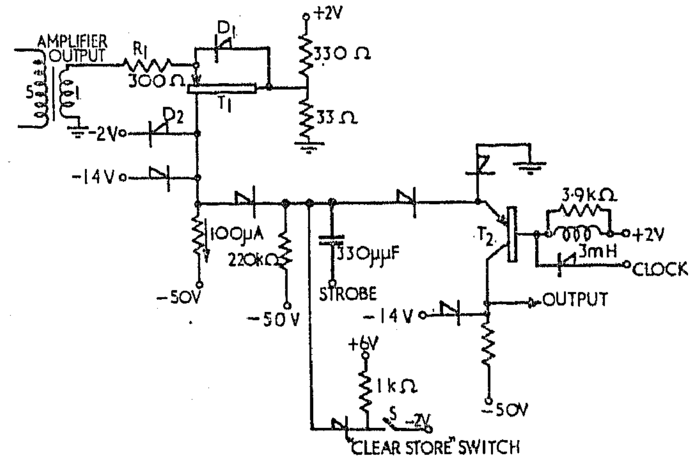


Fig. 26.—Squaring and retiming stages.

(9.4) Switching the Delay

In the processes of setting up numbers, and of multiplication, intermediate access and a switched change of delay are required. Both requirements are very easily satisfied with magnetostrictive stores by placing extra sending coils along the delay element as required. Only one extra transistor is needed for each sending coil, compared with the five required for each receiving coil. Intermediate coils do not affect the transmission properties of the line.

The switching of delays is accomplished by gating on the sending transistors.

(9.5) The Set-up Store

Intermediate access is necessary in the set-up store (Fig. 16) for the reasons mentioned in Section 7. Referring to the block diagram of Fig. 27, the two entry points to the line are constituted by

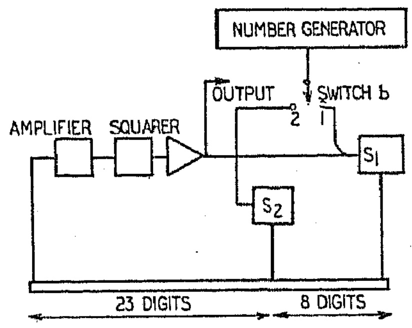


Fig. 27.—Set-up store.

the two sending transistors S_1 and S_2, separated by 8 digit periods and connected to the number generator by switch *b*. When the number generator is connected to S_1 the number will circulate in the normal way, S_2 being inoperative. When connected to S_2, however, S_1 will provide the regeneration, so that if the switch is closed for more than one circulation time S_2 will be putting digits into its coil at the same time as the same digits arrive under this coil from S_1. This results in a double-amplitude signal at the receiving coil, and the amplifier must be designed to cope with it.

(9.6) The *xy* Store.

The *xy* store (Fig. 28) is provided with two sending transistors, S_1 and S_2, having gating waveforms X and Y on their emitters so that they circulate numbers *x* and *y* respectively. The circulation time of *x* is thus 64 digit periods and that of *y* is 72 digit periods.

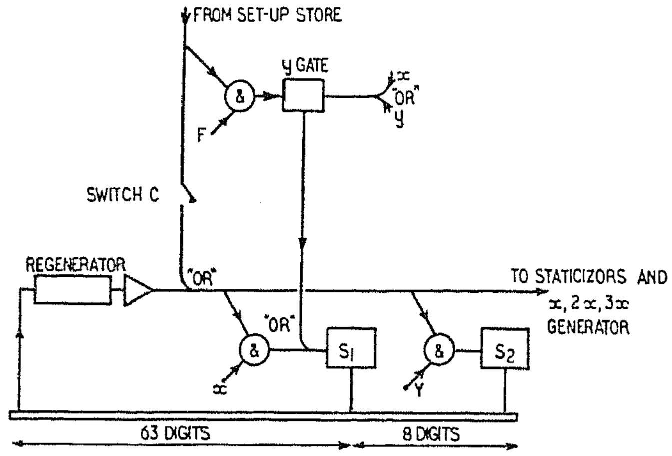


Fig. 28.—xy store with transfer system from set-up store.

The number x is transferred from the set-up store by switch c and circulates continuously through S_1. The number y is transferred automatically by the y gate at the start of multiplication in a "once only" operation by the gating waveform F on the emitter and "X or Y" on the base (see Fig. 14). It is thus initially injected by S_1 but subsequently circulated by S_2, suffering an extra 8-digit-period delay with respect to x each circulation time. By this means the four quarters of y are made to appear in turn at the output of the store at time Y_4 in the order y_D, y_C, y_B, y_A so that they can be gated out by Y_4 for setting up the staticizors (Section 8). After four circulation times the four quarters of y have all been gated out and y has been lost, although x remains circulating (Fig. 18).

(9.7) The Product Store

The product store (Fig. 17) has a 64-digit delay, composed of a 63-digit line followed by a single transistor delay. This allows the product to be tapped off one digit period from the end to compensate for the extra delay suffered by x in the x, $2x$ and $3x$ generator.

During multiplication it has the 8-digit delay of the adder chain in series with it, and reference to Fig. 18 will show that the resulting 72-digit delay gives the correct timing for addition of the four partial products.

Gates A and B are controlled by waveform F, and, although they operate in less than a digit period, there are, in fact, eight digit periods available to accomplish this switching.

(10) NEGATIVE NUMBERS

So far, the numbers x and y have both been assumed positive, but means must now be discussed for dealing with negative numbers.

A well-established method for representing negative numbers[6,7] is by means of complements in which the positive number, say x, is subtracted from a larger fixed positive number to form the complement $C(x)$. This complement has the characteristic that its most significant digit is always a 1, which therefore provides the means for distinguishing it from the original positive number, the most significant digit of which is always made 0. Thus, if x contains n digits having the values 2^0 to 2^{n-1}, then since the sum of all these is less than 2^n, its complement can be formed by subtracting it from 2^n, forming $C(x) = 2^n - x$. Since 2^n is merely a 1 in the nth place it can be ignored because the number x only extends to the $(n-1)$th place. It follows, therefore, that $C(x) = -x$. To convert back to a positive number it is only necessary to recomplement, since $C[C(x)] = 2^n - (2^n - x) = x$, the original number. The rule for complementing a number is, starting from the least significant digit, to leave it unchanged until after the first 1, and from then on to interchange 1's and 0's.

There are two methods of using the complement system in a multiplier: first, to design the arithmetic unit to deal with complements, the product itself appearing as a complement if negative, and secondly, to store the signs of the two inputs, convert them to positive numbers, multiply as for positive numbers, and then complement the product if the signs of the input numbers indicate that it should be negative. The first method does not always give the correct answer, and although simple correcting circuits can be devised to remedy this they increase the multiplication time. The second method has therefore been used.

(10.1) The Input Complementer

(10.1.1) **Logic.**

The input complementer is included in the xy store; the complete block diagram with the original store in heavy lines is shown in Fig. 29. The store is initially clear, and x is the first number

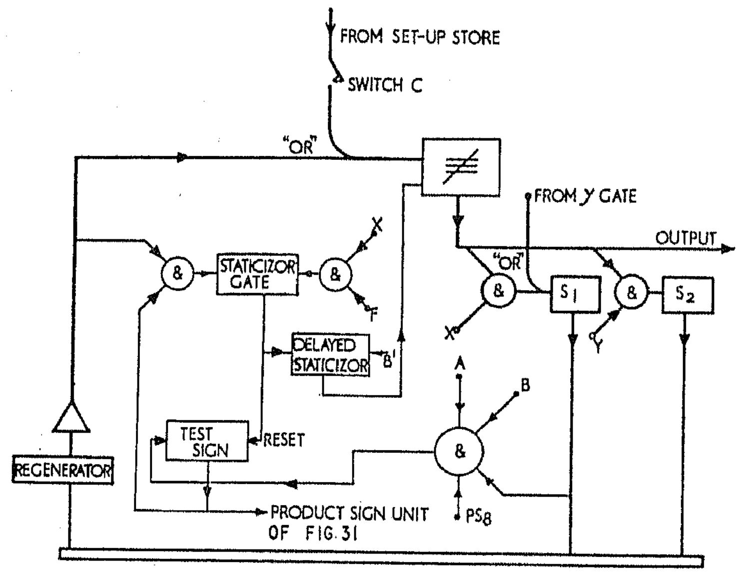


Fig. 29.—xy store, with complementing system shown in thin lines.

to be transferred from the set-up store. When it appears at the output of S_1, the "and" gate of A, B and PS_8 tests its most significant digit, which, if a 1, switches on the "test sign" staticizor. Assuming x is negative, on emerging from the delay line the first 1 to appear switches on the "staticizor gate," which in turn resets the test sign and also triggers the "delayed staticizor." The delayed staticizor waits for one digit period and then emits a row of 1's. The effect of this row of 1's on the "not-equivalent-to" circuit is to interchange 1's and 0's in the number (x) which was previously passing through it unchanged. The resulting recomplement of x is then circulated by S_1. Since its most significant digit is now 0, and the delayed staticizor has been reset by B′, no further complementing takes place.

The number y is fed into the xy store when multiplication begins and appears at the output of S_1, where its sign is tested as with x. At this moment, however, x is just emerging from the delay line, and, to prevent it being complemented in response to the sign digit of y, inhibits X and F are applied to the staticizor gate. The number y is complemented, if necessary, on emerging from the delay line and thereafter is circulated by S_2 as explained in Section 9.6.

(10.1.2) **Circuits.**

Each of the four blocks comprising the complementer uses one transistor, and since they are interconnected it is convenient to have them all at the same low voltage level.

The test sign transistor is the basic circuit of Fig. 4 with the "and" gate of A, B and PS_8 on its emitter and the transformer-reset of Fig. 5(*a*) on its base. The staticizor gate has gating on both emitter and base and is reset by its base coil after one or two digit periods.

The delayed staticizor circuit is shown in Fig. 30. Basically

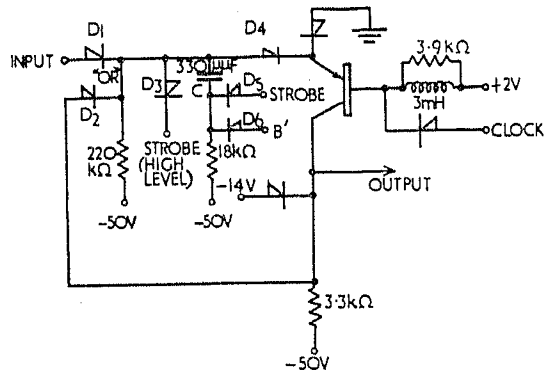


Fig. 30.—Delayed staticizor.

it is the delay stage of Fig. 8 but has an additional diode, D_2, connected from the collector to the input circuit. An input pulse on the anode of D_1 will cause an output pulse to be produced in the next digit period. This output pulse is then fed back through D_2 and provides the next input pulse, and so on, so that the circuit emits a continuous row of 1's. To reset the circuit the memory capacitor C must be discharged, and this is accomplished by the waveforms on the cathode of D_3 and the anode of D_6. The cathode of D_3 is clamped to earth potential between strobe pulses by a high-level version of the strobe, which has no effect on the operation of the circuit until a B′ pulse arrives on the anode of D_6. This B′ pulse occurs between two adjacent strobe pulses and takes the anode just above earth potential, which discharges the capacitor and destroys its memory since its other plate is simultaneously clamped at earth potential by D_3.

The "not-equivalent-to" circuit in Fig. 29 is identical to T_1 in the adder of Fig. 21.

(10.2) The Product Sign

The sign of the product is determined by whether the signs of the inputs are like or unlike, and this information is supplied by the circuit shown in Fig. 31.

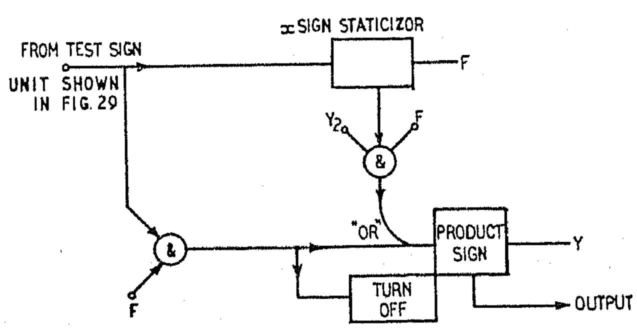


Fig. 31.—Block diagram of product-sign unit.

Since one of the numbers, x, may be retained for repeated multiplication, its sign must be remembered, and this is accomplished by the x-sign staticizor. Its base circuit consists of a current supply and catching diode, and it is triggered on by the sign digit of x, if it is negative, but not by y, since F inhibits it. When multiplication begins, a single sample of the sign of x is obtained by "(x sign) and Y_2 and F not Y," which switches on the product-sign transistor if x is negative. The product-sign transistor, together with its turn-off transistor, comprises a halver similar to that of Fig. 10 in that both transistors are mutually inhibiting. The sign of y is applied directly to this halver, and if y is negative the state of the halver is changed. The "and" gate, controlled by F, prevents x being fed directly to the halver. The output waveform of the product-sign circuit is thus positive if the input signs are different and negative if they are the same.

No provision is made for actually complementing the product, since it is assumed that the external computer has facilities for either adding or subtracting the product into its accumulator, depending on whether the product-sign output is at negative or positive potential.

(11) CONSTRUCTION AND RESULTS

No attempt was made to miniaturize the equipment, the circuits being constructed on tag boards 20 in long by 2·5 in wide mounted on a Post Office type relay rack. The three magnetostrictive delay lines are mounted vertically on a separate rack together with their amplifiers.

The complete multiplier comprises 96 transistors, 568 crystal diodes, 74 coils, 373 resistors and 117 capacitors. The ratio of crystal diodes to transistors (about 6) is roughly the ratio of their cost. The currents supplied by the plus and minus 50-volt lines are 200 mA and 760 mA respectively, resulting in a total power consumption of approximately 50 watts, an average of about 0·5 watt per transistor.

To test the multiplier some specimen multiplications of signed 32-digit numbers were tried on both the multiplier and a Ferranti computer. The results were in agreement in each case.

During seven months of daily operation, only three transistors, all point-contacts, have failed. Two of these failures took the form of a collapsed output characteristic, which may have been caused by overloading although no evidence of this was found. The third transistor acquired an i_{co} of 3 mA, but functioned perfectly when transferred to a circuit incorporating a base coil. Although only one such fault has occurred in the multiplier, it has been found that transistors left on the shelf are prone to this form of deterioration.

The wiring of the transistor circuits has proved to be extremely simple. This is primarily due to the absence of heaters and valve bases, but also arises partly from the low output impedance of the transistor regenerative amplifier, which enables stray capacitances largely to be ignored and the equivalent of the cathode-follower in valve circuits to be dispensed with.

(12) CONCLUSIONS

The use of powerful clock waveforms, generated by valves, simplifies design problems and allows the transistors to be operated well below their maximum power rating. It is thought that the low transistor failure rate is largely the result of under-running them.

The results also appear to have justified the use of base coils. If the operations of the circuits of Figs. 2 and 4 are compared, that of Fig. 4 is found to require a smaller triggering current and is also reset more rapidly. In addition to this, the circuit which has the base coil is virtually unaffected by any increase in i_{c0} caused by deterioration on the shelf, by temperature or by the residual effects of hole storage when in operation. In fact, transistors of either LS737 or OC51 type will operate equally well in any circuit in the multiplier which embodies a base coil, without selection, and this in spite of the relatively high operating frequency of 125 kc/s.

Although no estimate of long-term reliability can be made on the basis of seven months' operation, the short-term reliability of the point-contact transistor appears to be very good, and it is thought that this reliability, coupled with its versatility and

adequate frequency response, enables a good case to be made out for the continued manufacture of such transistors.

(13) ACKNOWLEDGMENTS

The authors are indebted to Dr. T. Kilburn for his guidance throughout the work described, and to Dr. A. A. Robinson and Mr. C. G. Scarrat, both of Ferranti Ltd., for helpful discussions.

(14) REFERENCES

(1) CHAPLIN, G. B. B.: "The Transistor Regenerative Amplifier as a Computer Element," *Proceedings I.E.E.*, Paper No. 1647 R, September, 1954 (**101**, Part III, p. 298).
(2) BASSET, H. G., and TILMAN, J. R.: "Some Transient Properties of Transistors," *British Journal of Applied Physics*, 1953, **4**, p. 116.
(3) FLORIDA, C. D., HOLT, F. R., and STEPHEN, J. H.: "Irradiation of Transistors," *Nature*, 1954, **173**, p. 397.
(4) NICHOLSON, A. W.: "High Speed Multiplier" (M.Sc. Thesis, University of Manchester, 1953).
(5) COOKE-YARBOROUGH, E. H.: "A Versatile Transistor Circuit," *Proceedings I.E.E.*, Paper No. 1613 R, September, 1954 (**101** III, p. 281).
(6) ROBINSON, A. A.: "Multiplication in the Manchester University High-Speed Digital Computer," *Electronic Engineering*, 1953, **25**, p. 6.
(7) SHAW, R. F.: "Arithmetic Operations in a Binary Computer," *Review of Scientific Instruments*, 1950, **21**, p. 687.
(8) ROBBINS, R. C., and MILLERSHIP, R.: "Applications of Magnetostriction Delay Lines" (Symposium on Automatic Digital Computation, N.P.L., March, 1953).
(9) BRADBURN, E. M.: "Magnetostrictive Delay Line," *Electrical Communication*, 1951, **28**, p. 46.
(10) WILLIAMS, F. C., and CHAPLIN, G. B. B.: "A Method of Designing Transistor Trigger Circuits," *Proceedings I.E.E.*, Paper No. 1428 R, July, 1953 (**100**, Part III, p. 228).

(15) APPENDICES

(15.1) Magnetostriction

It has been shown by Joule, Villari and others that certain materials exhibit a change in length upon application of a magnetizing force, and that the reverse effect is also present, namely that a stress applied to such a material produces a change in its magnetic state. If a polarizing force is applied in such a case, the change in magnetic state may be detected by the change in induction in the material.

(15.2) Application to a Delay Line

The effects of magnetostriction may be utilized in the construction of a delay line. The delay is obtained by converting electrical pulses into magnetic pulses, and thus into stress pulses, which are propagated at the speed of sound in the material, and are then detected by means of the reverse magnetostrictive effect and reconverted into electrical variations. The delay is entirely dependent upon the length and physical properties of the material. A practical arrangement is shown in Fig. 22. Two coils are wound longitudinally around a wire made of nickel, which is strongly magnetostrictive. If a step input of current is considered [Fig. 32(*a*)] it can be seen that a square stress pulse of physical length l_1 is formed under the coil. This is propagated along the wire and appears as a square pulse (*b*) of duration l_1/c under any single turn of the receiving coil. This produces a change in induction in the coil corresponding to the shape of the pulse, which is differentiated and appears as a voltage (*c*) in the coil. Integrating over all the turns in the coil gives the waveform (*d*). By inverting these waveforms the effect due to the negative-going edge of a square pulse of input current may be obtained, and added to Fig. 32, as shown by the dotted lines.

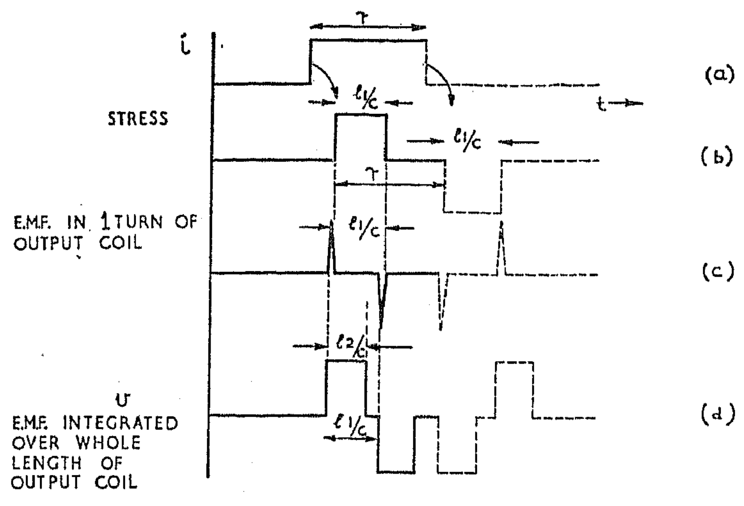


Fig. 32.—Magnetostrictive delay-line waveforms.

By making τ (the input pulse width) equal to l_1/c, the negative pulses may be made to superimpose, giving double-amplitude negative pulses. Also, by making $l_1 = l_2$, the time wasted between positive and negative parts of the output pulses is eliminated.

In practice, the polarity of the output pulse is reversed and the positive part detected [Fig. 33(*a*)].

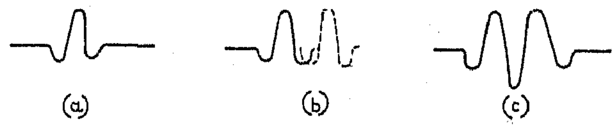


Fig. 33.—Optimum timing of adjacent pulses.
(*a*) Single pulse.
(*b*) Positioning of adjacent pulses for optimum timing.
(*c*) Resultant waveform.

It should be noted that pulses may be packed so that their negative parts overlap (*b*) without any loss in height of the positive portions (*c*).

(15.3) Design of the Delay Line

In order to increase the efficiency of the delay line a stranded form of construction was used. Owing to skin effect, the outer portions only of the wire are magnetized at the frequency of operation. A larger surface area obtained by the use of wire of heavier gauge would be advantageous, but might give rise to unwanted modes of propagation if the wire diameter approached the dimensions of the component wavelengths of the stress. It is desirable, therefore, to keep the diameter as small as possible and to obtain the necessary surface area by using a stranded wire.

It was decided to use 100 strands of 0·004 in-diameter nickel wire as satisfying the above conditions. Above 100 strands the gain in output tends to fall below the theoretical value, and the line tends to become physically unmanageable.

Stress pulses are propagated best in hard-drawn material, since there is little dispersion and the attenuation is slight. The magnetostrictive effects, however, are stronger in annealed material by a factor of nearly 1·7. To combine the advantages of both states, hard-drawn material is used and the portions under the coils are annealed. This gives a total increase in output of three times that in the hard-drawn state.

An acoustic delay line is similar to an electric one in that reflections occur at the ends. To overcome this, the pulses approaching the ends of the wire are passed through synthetic

rubber damping pads clamped about the wire. These pads produce considerable attenuation of the stress pulses, and by the time the reflected pulses have passed through the pads a second time they are reduced to negligible proportions.

To aid the suppression of echoes, the portions of line between each coil and the end are annealed, producing further attenuation. A certain cancellation of echo is also obtained by clamping half the strands at the ends and leaving the other half free. The echoes produced at these terminations are of opposite sign and tend to cancel out when passing through the coil.

The coils must have their lengths accurately determined to match the input pulse, and they should also be highly efficient magnetically, since low levels of current and voltage are to be used. They are therefore fitted with ferrite core materials to direct all the flux into the wire, and also to define the edges of the field exactly.

DISCUSSION ON

"DETERMINATION OF THE STATIC AND DYNAMIC ELASTIC PROPERTIES OF RESILIENT MATERIALS"*

NORTH-WESTERN MEASUREMENTS GROUP, AT MANCHESTER, 23RD NOVEMBER, 1954, AND NORTH-EASTERN RADIO AND MEASUREMENTS GROUP, AT NEWCASTLE UPON TYNE, 17TH JANUARY, 1955

Mr. F. W. Taylor (*at Manchester*): Since everyone is becoming more and more conscious of the noise emitted by electrical plant and machinery generally, manufacturers are being compelled to give serious consideration to new materials and techniques for its reduction. The old methods are no longer satisfactory, and a much closer study of the fundamentals involved, such as the material characteristics, is essential.

I agree that there is a distinct lack of data suitable for use by the engineer called upon to solve these noise problems. The makers of sound- or vibration-absorbing materials have so far only published static characteristics. Consequently, users are compelled either to base calculations on these, very often with disappointing results, or, if they are sufficiently knowledgeable, to make a more or less intelligent guess at the dynamic properties.

In their previous work the authors have warned of the dangers of using the wrong figures, and the results of the tests outlined in the paper show how misleading calculations based on the static figures could have been. For example, assuming a frequency of 100 c/s, which is generally the fundamental frequency for most magnetically generated noise and vibrations, the ratio of the dynamic to the static stiffness is about 2 for soft rubber, about 3 to 4 for the synthetic rubber tested and over 6 for cork when the static loading is only 60 lb/in^2. With other materials the ratio is even higher.

I would therefore ask that manufacturers of resilient and absorbent materials make the fullest possible use of instruments and test gear such as the authors describe, and so obtain a proper appreciation of the effectiveness of their products. This should enable them to develop new and better materials and to supply their customers with more practical information. Also, since transformers and electrical apparatus generally are expected to remain in position for many years, it would be an advantage if the makers would give some idea of the creep properties and also of how the static and dynamic characteristics are likely to vary over the years.

It is essential that the framework, which is not given much attention in the paper, should be exceedingly rigid, otherwise misleading results are obtained.

When any material is being tested, the question of the repeatability of the results and representative sampling always arises. To what degree of accuracy can the dynamic stiffness be measured? How consistent are the results on any given sample, and what variations in properties of different samples of the same material were encountered in the materials covered in the paper?

In Figs. 5–10 special attempts were made, of course, to keep the waveform of the driving force pure. Since most noise consists of a fundamental and several, if not many, harmonics, would it not be possible to test the samples with a complicated waveform and so get a more useful figure for the dynamic stiffness?

All the tests mentioned in the paper were carried out in air. In a transformer, however, where the core is mounted in the tank on pads of resilient material, the pads are surrounded by oil, which is really in parallel with the resilient material. Would the oil invalidate any calculations based on the properties obtained with the machines? In other words, how would the performance of the material be affected by the oil? There is no doubt that the effect of pads of a given material differs according to whether the transformer tank is filled with air or oil.

Mr. G. H. Hickling (*at Newcastle upon Tyne*): Although my experience of resilient mountings has been much more restricted than that of the authors, I have occasionally employed rubber and similar supports, mainly for the purpose of protecting delicate equipment against outside sources of vibration. Therefore, tests have been made on various rubbers, felt, granulated cork, etc., using somewhat simpler measuring devices than those described in the paper.

I have used a test instrument for rubber-sheet samples essentially similar to that of Fig. 10 in Reference 2 of the paper, except that standard crystal-type accelerometers were used for displacement measurements, which were restricted to resonance conditions. It was deliberately arranged to test comparatively stiff flat-sheet specimens in order to determine the properties of the material independently of the shape effect due to bulging at the edges. Thus the test frequencies were necessarily high, and ranged, with three masses available, from 600 to 2 000 c/s. Under these conditions rubber of a Shore hardness of 70 gave extremely high ratios of dynamic to static stiffness, which exceeded 10 and which remained almost constant throughout the above frequency range.

For the purpose of studying the behaviour of rubber at lower frequencies (1–10 c/s), another very simple test equipment was used, comprising a heavy steel beam arranged to see-saw on a small roller and restored to the horizontal position by the pair of small rubber blocks under test, which were placed beneath it on either side of the fulcrum. Adjustable packing under these

* JACKSON, R. S., KING, A. J., and MAGUIRE, C. R.: Paper No. 1579 M, November, 1953 (see 101, Part II, p. 512).

blocks enabled the desired static compression to be obtained. The natural frequency of this system is readily adjustable over a wide range by varying the distance of the test blocks from the fulcrum. With this apparatus and with samples of more nearly cubic proportions, it was found that the dynamic/static stiffness ratio still ranged between 2 and 4, depending on the static loading, and tending towards the upper limit at the maximum frequency of 10 c/s.

The standing-wave, or self-resonance, measurements described in the paper are of special interest in relation to noise transmission. The statement in Section 8.2 implies that the transmitted force factor of approximately 7 results principally from the damping of the material, i.e. from the magnification factor of the resonance curves (see Fig. 12). The low-frequency curves, however, all show increasing stiffness with frequency. To what extent, therefore, can this increased force transmission be attributed to the properties of the rubber itself as distinct from resonance effects?

Mr. H. J. Bernhard (*at Newcastle upon Tyne*): A good deal of work has been expended on the design and development of the apparatus used for these rather important tests, and I only wish that the results of more tests could have been given and trust that further experiments will be made with the apparatus.

An instrument was developed on exactly the same principles by Costadoni* in 1936. I do not recollect the dimensions of the specimens he investigated, but some of his test results agreed very well with those obtained by the authors. He found dynamic stiffnesses up to 20 times higher than the static stiffness. He also mentioned that there is a certain static load for average rubber and cork which gives the best damping effect. The load is about 100 lb/in^2, and it agrees very well with Fig. 8(*b*), although not so well with Fig. 9(*b*).

The dynamic tests shown in Figs. 5(*b*), 6(*b*), 7(*b*), 8(*b*), 9(*b*) and 10(*b*) were obviously made at 100 c/s. Have the authors also made tests with varying static loads at other frequencies, and would they then expect the curves to be the same? If not, I consider that such tests should be made, because it would obviously be wrong to use values obtained at 100 c/s for very different frequencies.

Do the dynamic stiffness and magnification depend on the shape of the test specimens, or are the values found in the experiments material constants which are independent of the shape? Many so-called material constants do, in fact, vary with the dimensions of the test specimen, e.g. elongation of steel, Izod values, fatigue strength, etc. As nobody would want to stress a bolt to give an elongation of 30%, the variation in elongation with length of steel is perhaps not so bad, and such characteristics are generally used to give an indication of other properties of the material, such as ductility. However, if a system is designed with a certain natural frequency it is essential that the true value for stiffness be known and used. This may vary considerably between, for example, 100 in^2 of $\frac{1}{2}$ in-thick rubber sheet and a cube of 5 in side. Although large specimens cannot be accommodated in the authors' apparatus, the thickness of test pieces could be varied, and tests with varying thicknesses could be treated as model tests.

* Costadoni, C.: "Electrodynamic Machine for Testing Damping Materials," *Zeitschrift für technische Physik*, 1936, 17, p. 108.

Mr. E. Meadley (*at Newcastle upon Tyne*): The authors mention "soft" rubber sheet and then quote values of 50–55 for Shore hardness, or even 69 (see Fig. 14). In some applications soft rubber is usually assumed to have values of 35–40. Perhaps the authors would discuss the effect of Shore hardness on stiffness. In particular, is the value of 100 lb/in^2 for static loading dependent upon the Shore hardness of the sample?

Mr. R. S. Jackson, Dr. A. J. King and **Mr. C. R. Maguire** (*in reply*): We heartily endorse Mr. F. W. Taylor's plea for more quantitative information from the manufacturers of resilient materials and vibration-attenuating supports on the performance of their products with load, frequency and time. The apparatus described can determine these elastic properties to within 3%, which is also the order of repeatability of tests on samples cut from the same piece of material. Tests on bonded samples of rubber are better in this respect than those on unbonded samples owing to the more definite conditions at the loading surfaces. Tests with a complex waveform, as opposed to a single frequency, would be of little use, since the attenuation of a mounting varies with frequency as well as stiffness.

In oil two questions arise: the ability of the resilient material to withstand oil and the transmission of vibration through the oil. Oil-resisting rubbers are available but will provide little net attenuation of vibration between a transformer core and tank unless transmission through the oil is also reduced.

Mr. Hickling's experiences in testing rubber sheets are very interesting, especially the high values of dynamic/static ratio which he found. The stiffness of a piece of rubber or other material at its first thickness mode is determined, like that of any parallel-resonant system, by the Q-factor of the system, and therefore, as the losses take place in the material, by the damping properties or Q-factor of the material. The trough between the first and second thickness modes seen in Fig. 12 (*b*) gives some indication of the relative contributions of stiffness and resonance to the transmission, but the factor 7 was, as stated in the paper, based on low-frequency data only.

A reference to Costadoni's paper, quoted by Mr. Bernhard, was given in a previous paper by one of the authors (Reference 2 in the paper).

The dynamic tests shown in Figs. 5(*b*) to 10(*b*) were made at 60 c/s. For each material, measurements were made over the entire frequency range for three different static loads, and the shapes of the curves were so similar to those presented that it is permissible to use curves (*a*) to represent the variations with frequency, and curves (*b*) to act as "scaling factors."

Some information on the effect of shape factor is given in our Reference 1.

We have never had occasion to test the very soft rubber quoted by Mr. Meadley as of Shore hardness 35–40. The paper is concerned with the apparatus and method of test, so that only typical test results on common resilient materials have been included. Many more tests on rubbers with a range of Shore hardness would be required to determine the corresponding optimum loadings.

621.317.3/.4

Paper No. 1703 M
Sept. 1954

ELECTRICAL AND MAGNETIC MEASUREMENTS IN AN ELECTRICAL ENGINEERING FACTORY

By D. EDMUNDSON, B.Sc.(Eng.), Member.

(*The paper was first received* 27*th February, and in revised form* 29*th April*, 1954. *It was published in September*, 1954, *and was read before the* MEASUREMENTS SECTION 1*st February, the* NORTH-WESTERN MEASUREMENTS GROUP 22*nd February, and the* NORTH STAFFORDSHIRE SUB-CENTRE 3*rd May*, 1955.)

SUMMARY

The paper describes how, by the application of negative feedback amplifiers and high-speed polarized relays, new methods of performing the commoner electrical and magnetic measurements in power engineering have been developed. Examples include the application of amplifier techniques to the measurement of power and voltage, and of high-speed relays to the measurement of speed, frequency and the slip frequency of induction motors. In a Section on industrial magnetic measurements, apparatus is described for routine tests on whole sheets of transformer steel and on toroidal specimens; while a new approach to routine permeability measurements is given. The paper concludes with an example in the field of dielectric measurements—the rapid measurement of the power-factor of insulation samples.

LIST OF PRINCIPAL SYMBOLS

Since the paper describes a variety of measuring instruments, the symbols are, in general, those in normal use for each purpose, and are explained in each Section. In Section 4, however, which deals with magnetic measurements, the following symbols have been used throughout:

Φ = Instantaneous value of total flux, maxwells.
A = Cross-sectional area of magnetic path, cm^2.
B = Instantaneous value of Φ/A, or apparent flux-density, gauss.
$\hat{B}$ = Maximum value of B.
H = Instantaneous value of magnetizing force at specimen surface, oersteds.
$\hat{H}$ = Maximum value of H.

(1) INTRODUCTION

The paper has been written with the object of surveying the field of electrical measurements in a single industrial organization, in order to record the more important developments which have occurred during the period beginning about 1939.

The apparatus described is in no instance of a research nature. Every example has been developed for, and by, the testing organization of a large electrical manufacturing concern. Hence the emphasis throughout is practical. Several of the instruments are used by unskilled operators, and all are in regular operation, mostly under factory conditions.

(1.1) Scope and Omissions

The paper is restricted to the more common measurements in power engineering. It omits a number of subjects, such as strain-gauge, temperature and vibration measurements—which have been exhaustively treated elsewhere—and special apparatus which is not of general interest or application.

(1.2) General Approach

The most important single change which has come about during the period under review has been the introduction of electronic techniques, which before 1940 were regarded among measurements engineers with universal suspicion. One of two conditions has been observed in all applications. In handling measurements of frequency or speed, amplifier distortion is seldom harmful, for the form, amplitude or phase of the output is usually immaterial, while no amplifier, however non-linear, is likely to alter the frequency of the signal. In all other measurements, sufficient negative feedback has been included to eliminate any question of the measurement being dependent on valve characteristics or supply voltage.

One other component which has been widely used, but which is not a feature of textbooks on electrical measurements, is the polarized relay. After considerable use of various home-made designs, the author has found the Carpenter relay, when correctly used, ideal for all his applications. Although referred to briefly by Casson and Last,[1] little has been written about the convenience of a sensitive relay which will operate at frequencies up to 500 c/s and act as a mechanical rectifier, inverter or condenser charger with equal success. Its availability has opened up as many fields of measurement as the valve amplifier itself; its versatility is well illustrated in the course of the paper.

Perhaps the most striking contrast with measurement innovations of the past, however, lies in the tendency to construct new measuring apparatus of standard components rather than to develop essentially new instruments. Little demand, in the subject-matter of the paper, is made on the skill of the instrument maker.

(2) MEASUREMENTS OF ELECTRICAL QUANTITIES AT POWER FREQUENCIES

Current, voltage and power are the three quantities whose measurement is common to the whole field of electrical engineering, and for a long time highly accurate a.c. instruments and instrument transformers have been available. There is, however, a wide field where difficulties arise in their use even at power frequencies. An essential condition of all measurement work is that the introduction of measuring apparatus must not disturb the system being measured. This is equally true, say, of temperature or speed: the power consumption of the instruments must be negligible compared with that of the quantity being measured. This condition cannot be met when measuring the consumption of small motors, lamps and so on; nor is it usually possible to wind wattmeters for the low voltages frequently encountered.

These difficulties can be completely overcome by the use of an amplifier to supply the voltmeter and the voltage coil of the wattmeter. Careful design ensures that the accuracy is not diminished, but rather the reverse—for a series resistor is not an ideal device for regulating the current in a wattmeter moving-coil. The real advantage of the amplifier for this purpose, apart from its high-impedance input, lies in the fact that the "load"—i.e. the instrument—has a negligible influence on the current passing.

(2.1) Standard Wattmeter Amplifier

Amplifier wattmeters have been designed in the past,[2, 3] but have not found general acceptance. The circuit shown in Fig. 1

Mr. Edmundson is with the British Thomson-Houston Co., Ltd.

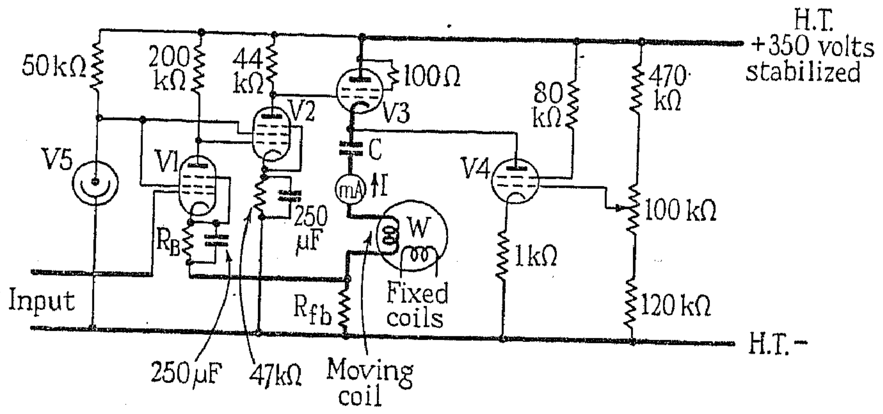


Fig. 1.—Wattmeter amplifier.

is of universal application, although first used for magnetic measurements. It might be described as an amplified cathode-follower. The wattmeter voltage coil and the feedback resistance are connected in the cathode circuit of V3, whose grid is supplied by a 2-stage direct-coupled amplifier, the input to which is the difference between the input voltage and the voltage drop across the feedback resistor, R_{fb}. A condenser diverts the d.c. component of the cathode current of V3 through a tetrode, V4, connected in such a way that its a.c. resistance is enhanced. In effect V4 acts as a high-efficiency choke without phase-shift.

The input to V1, which is the error, is usually about 1 mV for an output current of 20 mA (the current for which commercial-wattmeter moving-coils are normally designed). If, therefore, the amplifier is designed for an input of 1 volt or more—or in other words, if the value of the feedback resistor is at least 50 ohms—the errors whether of ratio or phase do not exceed 0·1%.

The amplifier can therefore be designed for inputs of 1 volt upwards, and its load on a circuit is negligible. For inputs greater than 10 volts it is more convenient to provide an input voltage divider, but this also can be of such high impedance that its consumption can be neglected: 10 kilohms per volt is a convenient value. The circuit has one additional advantage: it saturates at an overload which is too small to damage the meter coil. Thus the circuit provides automatic protection.

(2.2) Amplifier Performance and Accuracy

The simplest method of checking the performance of the circuit is shown in Fig. 2. In a wattmeter with a 5-amp current

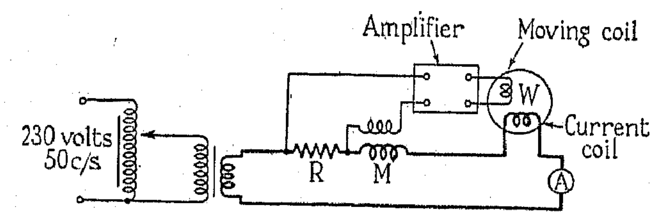


Fig. 2.—Calibrating circuit for amplifier wattmeter.

coil, and an amplifier designed for a 5-volt input, appropriate values for R and M would be 1 ohm and 3·18 mH respectively. If R is short-circuited, the mutual inductance applies the designed input voltage in exact quadrature, and although both coils carry their rated currents, the instrument should not deflect. If the primary winding of M is short-circuited, the instrument reads the power lost in R, which can be accurately known. By the application of this test, an instrument designed for 50-c/s use shows no readable phase-angle error up to 2 000 c/s, nor ratio error up to 5 000 c/s. A reduction in the value of C (Fig. 1) can provide for higher frequency ranges if desired. The performance is unaffected either by valve exchange or by supply voltage variations. The stabilized power supply is included, not because the amplifier is sensitive to h.t. voltage variations, but because a low anode-supply impedance is essential.

(2.3) Application to Light-Power Measurements

The most immediate use of the amplifier wattmeter lies in light-power measurements. For this purpose the circuit shown in Fig. 3 is being adopted as a universal measuring set. A

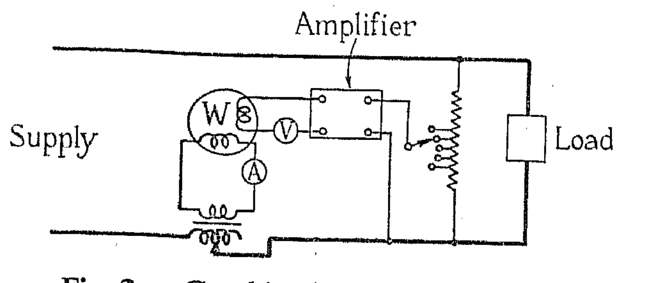


Fig. 3.—Combination measuring unit.

milliammeter in series with the wattmeter voltage coil is calibrated to read in volts. Any convenient ranges can be employed—for example:

Volts	1	$\sqrt{(10)}$	10	$10\sqrt{(10)}$	100	$100\sqrt{(10)}$	1 000.
Amperes	$\sqrt{(10)}/10$	1	$\sqrt{(10)}$	10	$10\sqrt{(10)}$.		
Watts	varying from 0·3 to 30 000.						

It will be noted that the high-resistance voltage divider which supplies the wattmeter voltage-coil is connected direct across the load, while the current ranges are obtained from a current transformer with a tapped primary winding.

(2.4) Measurements at Low Power-Factors

There is one particular field of measurements for which ordinary wattmeters are unsuitable even when the power is large. At low power-factors, the small deflection with rated current limits the accuracy of reading. A.C. bridges, although widely used where an impedance measurement only is required, are inconvenient when a load has to be set to a definite value (as in short-circuit tests on transformers), and not well adapted for either ferromagnetic or polyphase work.

For such applications the amplifier wattmeter, used with a backing-off mutual inductance, is highly convenient. Fig. 4

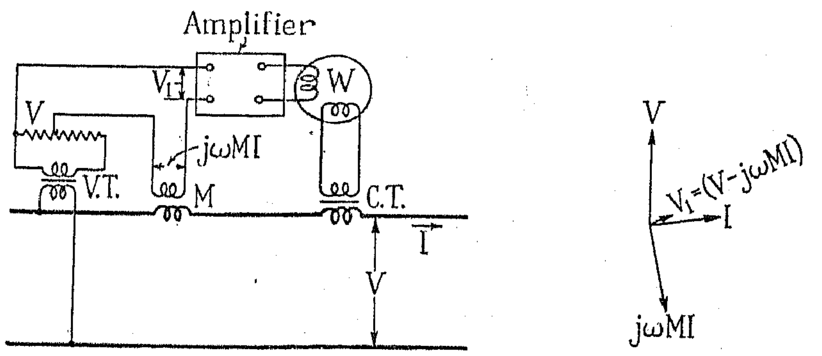


Fig. 4.—Low-power-factor measurement.

shows the connections for measurement of a substantial load where instrument transformers have to be used, together with a vector diagram in which, for simplicity, transformer and voltage-divider ratios have been omitted. V_1, the resultant input to the amplifier, has been shown slightly out of phase with I, since the exact value of M is immaterial. Since the wattmeter is now working near unity power-factor a substantial deflection can be obtained. The use of a toroidal mutual-inductance in the primary circuit virtually eliminates phase-angle errors in the current transformer; while the voltage transformer, working on open-circuit, can be designed with very low errors indeed.

Although drawn for single-phase use, this method can be

readily applied to the 2-wattmeter method of power measurement, using a polyphase wattmeter.

(3) MEASUREMENT OF SPEED AND FREQUENCY

Speed, or angular velocity, is perhaps the most important measurement which is common to mechanical and electrical engineering. Although any desired degree of accuracy is obtainable by chronometric methods, the use of these demands a period of time far greater than would normally be acceptable if it were considered as the period of an indicating instrument. For remote indication, electrical tachometers are common; but their accuracy is limited by that of the tacho-generator, which is a rotating machine, by a commutator, or by rectifiers.

If, however, an instrument is used to measure the frequency, rather than the voltage, of an a.c. tacho-generator, both it and the resistance of the leads to the instrument cease to have any effect on the accuracy of measurement. This was the consideration that gave rise to the development of a series of instruments, which, although in themselves frequency meters, are more often used for the measurement of rotational speed.

The instruments to be described all derive from Maxwell's commutator bridge. Developed by Clerk Maxwell to measure in electromagnetic units the capacitance of a condenser whose capacitance in electrostatic units had been determined by mensuration, in terms of known resistances and a constant value of frequency, the "commutator bridge" is here used with a known capacitance to measure an unknown frequency or speed. Although described by Maxwell[4] in almost non-mathematical terms with his usual clarity of language, simple in conception, capable of the utmost accuracy, convenient and versatile, the circuit has been much neglected. It has even been suggested[5] that it is not suitable for high accuracy without modification. These misconceptions are readily removed when Maxwell's original treatment is studied.

Before describing three instruments, selected as illustrating the versatility of the method, a short summary of its theory, common design features, and general advantages is given in Section 3.1.

(3.1) Theory of Maxwell's Commutator Bridge—Balanced and Unbalanced

Maxwell pointed out[4] that a succession of impulsive currents obtained by charging a condenser, C, n times a second from a battery, would produce a mean current in a moving-coil galvanometer of the same value as a resistor R,

where $R = \dfrac{1}{nC}$ − (Circuit resistance measured from condenser terminals)

He therefore placed the condenser and its charging device in one arm of a Wheatstone bridge: in Fig. 5 the condenser is

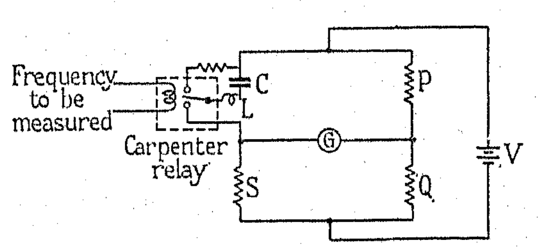


Fig. 5.—Maxwell's commutator bridge—elementary circuit.

discharged through an auxiliary resistor by a polarized relay, whereas Maxwell used a reversing commutator, but there is no difference in principle. The mean galvanometer current will now be zero when $R/S = P/Q$, using the symbols of Fig. 5; while, if the resistance of the battery is negligible and that of the galvanometer is G,

$$R = \frac{1}{nC} - \frac{\left(\dfrac{PQ}{P+Q} + G\right)S}{\dfrac{PQ}{P+Q} + G + S} \quad . \quad . \quad . \quad (1)$$

This expression is correct only if the condenser is allowed to charge completely at each operation of the polarized relay, a condition which can be met if the fraction on the right of eqn. (1), i.e. the circuit resistance, is small compared with $1/nC$. Its evaluation is therefore no more than a matter of mental arithmetic. The expression for bridge balance due to Sir J. J. Thompson,[6] which is commonly quoted but which is far less convenient, can be readily derived from eqn. (1) by algebraic processes.

If the components of the bridge are arranged to provide balance at some charging frequency n_0, then the galvanometer current at any other frequency is

$$i_G = V\frac{Q}{(G+S)(P+Q)+PQ}\left(1 - \frac{n}{n_0}\right) \quad . \quad . \quad (2)$$

assuming again that the battery resistance is negligible. The galvanometer current is linear in frequency, and proportional to the difference between the frequency being measured and some other arbitrary frequency.

(3.2) Method of Application and General Design Considerations

The circuit can be used in two ways. As a null device, the value of n can be measured with the same accuracy as that of a resistance: while from eqn. (2) it is seen that, as a deflectional instrument, linear in speed or frequency, any desired part of a range can be covered in exactly the same way as with a resistance thermometer. In either instance the accuracy depends on the use of precision resistors and a stable condenser, which must be selected with care not only for a minimum variation of capacitance with temperature, but a very small residual charge effect—i.e. a low power factor. There is no difficulty in reducing errors to less than 0·1%.

(3.2.1) Design of Relay Circuit.

The most convenient agent for charging the condenser is a polarized relay. Carpenter's relay, already referred to, is capable of prolonged operation at frequencies of several hundred cycles per second, its coil demanding a current of only a few milliamperes from the circuit being measured. The exact contact time is of no importance. It will be seen that the contact duty is light, since no current passes at the moment of "break." Nevertheless, there is a danger of destructive sparking at "make" when the bridge voltage exceeds about 30 volts with, say, a condenser of 1 μF. Even this can be obviated by the use of a series inductance (see L in Fig, 5) of such a value that, with the condenser, C, and the effective bridge resistance, critical damping is obtained. The current rise is now delayed until the contacts are "home," while its decay is even more rapid. With this simple precaution and a suitable discharge resistance, a condenser of 1 μF can be charged from a 200-volt supply, producing, at 100 c/s, a mean current of 20 mA.

There are occasions—of which the slip-meter (Section 3.5) is an example—when even the small exciting current of the relay coil is inadmissible. It is then a simple matter to interpose an amplifier which reduces the demand on the circuit to negligible amounts.

(3.2.2) **Design of Bridge Circuit.**

Considered as a Wheatstone bridge, the galvanometer sensitivity is bound to be low because of the requirement that either S or $(Q + G)$ must be small compared with R, in the interests of efficient charging. For maximum sensitivity, P is made approximately equal to S. As a deflectional multi-range instrument it is convenient, however, to make Q equal to S and adjust P for different ranges. As a null instrument, on the other hand, n is proportional to Q provided that the bridge resistance, as measured from the condenser terminals, is negligible. Although this condition is unattainable, it can be simulated by the use of a large condenser shunting S. If, then, $C = 10^{-6}$, and $P = S = 1\,000$, $n = Q$, so that a scale of ohms for Q is also a scale of cycles per second.

(3.2.3) **General Advantages.**

The general advantages may be summarized as follows:

(*a*) Since the instruments are essentially counting devices, their accuracy is unaffected by waveform or voltage.

(*b*) When used to measure speed, the instruments can be operated from any type of alternator, or equally from magnetic, electrostatic or photo-electric pick-ups, or mechanical contacts.

(*c*) The accuracy is dependent on the values of stable components, and since the range of a deflectional instrument can be as small as may be desired, reading accuracy can be made to match the inherent accuracy of the circuit.

(3.3) Mains Frequency-Meter

This instrument is made necessary by the mains frequency variations common since the 1939–45 War. In the circuit of Fig. 6 it is seen that not only the relay-coil supply but also the

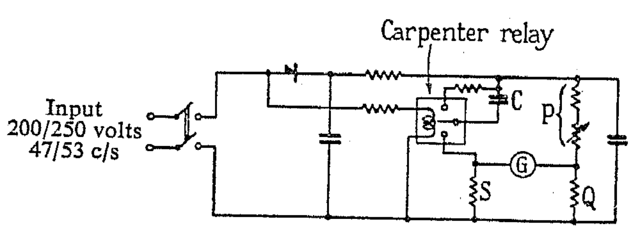


Fig. 6.—Mains frequency-meter.

bridge d.c. supply, derived by half-wave rectification, is obtained from the voltage whose frequency is to be measured. The circuit uses a pointer galvanometer. The bridge is designed to balance at any frequency within the range 47–53 c/s by adjustment of a small potentiometer, which forms parts of the resistor P, scaled in frequency. It is small, portable, convenient, and accurate to 0·05 c/s. A number of these instruments, made in 1945, have retained their accuracy unaltered up to the present time.

(3.4) Indicating Tachometer

Maxwell's commutator bridge can be used to operate an indicating instrument, provided either that the bridge voltage possesses short-term stability or that a ratiometer is used with one coil in place of the galvanometer and the other supplied through a fixed resistance from the same source as the bridge.

The first alternative includes a battery to supply the bridge and a rheostat in series with the galvanometer to adjust for battery-voltage variations. This adjustment resembles that of a resistance thermometer, in that it is made, before a reading is taken, under arbitrary conditions of known value. The most convenient arbitrary condition is zero speed, or frequency, which can be obtained at any time simply by open-circuiting the operating coil of the Carpenter relay.

Eqn. (2) shows that the galvanometer deflection is linear in speed, but that the scale of the instrument can cover any desired portion of the total speed range; for n_0 depends solely on the bridge components, and the maximum value of n/n_0 is limited only by the galvanometer sensitivity. Further, the range can be varied by adjusting the value of P (Fig. 5). A convenient arrangement is to provide a number of ranges such that the upper limit of one is the lower limit of the next: thus, an instrument reading up to 4 000 r.p.m. has ranges of 0–1 000, 1 000–2 000, 2 000–3 000 and 3 000–4 000. The effective scale length, and the scale accuracy of the instrument, are multiplied by the number of ranges.

(3.5) Slip Meter

A special problem of unusual interest which indicates the versatility of this technique is the direct measurement of the slip frequency of an induction motor. Although this quantity may be derived from the difference between actual and synchronous speeds, such a procedure is cumbersome; for each must be measured to a degree of accuracy greater than that actually required by a factor which is the inverse of the proportional slip. Thus to measure a slip of 5% with 1% accuracy requires speed and frequency measurement to an accuracy of 2·5 parts in 10^4.

The special difficulty of direct slip measurement lies in the very low frequencies concerned: a range of 0–7 c/s would cover ordinary 50-c/s motors up to their pull-out points. In this case the problem has been solved by the use of a thermal instrument whose naturally slow response has been turned to advantage. Referring to Fig. 7, the condenser C is charged by the Carpenter relay, and discharged into the vacuo-junction of a thermal milliammeter, once for every slip cycle. The voltage to which the condenser charges is established by the striking of a stabilizing tube.

In the classical theory of the Maxwell commutator bridge, it is an essential requirement that the torque acting on the galvanometer shall be strictly proportional to the instantaneous value of current passing through it, so that the mean torque shall be proportional to the integrated current over a discharge cycle, or total quantity of electricity stored in the condenser. The mean current passing the condenser is then

$$I_{mean} = nCV \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

But the alternative now adopted, of measuring the mean stored energy per cycle, is equally valid, and merely demands as an additional condition that the total resistance, R (Fig. 7), in the

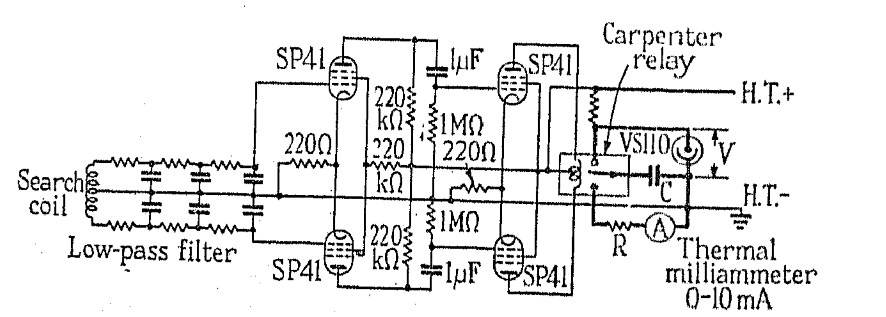


Fig. 7.—Slip-meter.

discharge circuit shall be known. The energy stored in the condenser must now be equal to that dissipated in R:

$$\tfrac{1}{2}nCV^2 = I^2_{\text{r.m.s.}}R \quad \text{or}$$

$$I_{\text{r.m.s.}} = \sqrt{\left(\frac{nC}{2R}\right)}V \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Since the scale of the thermal milliammeter is proportional to $I^2_{\text{r.m.s.}}$, it is linear in frequency. Moreover, the r.m.s. value of the train of impulses greatly exceeds the mean value. For instance, if V is 100 volts, C is 4 μF and R is 250 ohms, then

for 5 c/s the mean current is 2 mA, but the r.m.s. current is 20 mA. This enables an instrument to be used which is fairly robust, and whose thermal time-lag is long enough for a frequency of 1 c/s not to cause very much movement of the needle.

The rest of the circuit of Fig. 7 is concerned with producing a slip-frequency signal of sufficient magnitude to operate the polarized relay. Although in principle the slip-frequency flux is confined to the rotor of an induction motor, the inevitable slight variations in manufacture give rise to a small axial component which can be detected by a search coil placed near the end of the shaft. The search coil invariably picks up a much larger voltage at mains frequency; before the slip-frequency voltage can be amplified and made to operate the polarized relay this is attenuated by a simple 3-stage low-pass filter. A symmetrical amplifier facilitates the handling of the low-frequency signal.

(4) MAGNETIC MEASUREMENTS

Magnetic sheet and strip are the most valuable, critical and variable raw materials of the electrical industry. The measurement and control of their properties is an important task in a works laboratory. The use of techniques similar to those already described has made possible the development of apparatus which will perform measurements in industrial conditions which have previously required delicate laboratory methods.

Of all the soft magnetic material produced in this country, about 70% is in the form of low-silicon-content steel sheet or strip, used for small motors, for which accurate control of the properties is not essential. But for the steels and alloys used in transformers, and in certain classes of large alternator, the losses must be known with certainty before the material is punched into laminations; the characteristics of the core materials of instrument transformers must also be individually controlled.

In this Section are described, therefore, first, an apparatus for the grading of individual sheets of transformer steel according to their specific loss; and secondly, an apparatus for routine measurements on toroidal strip-wound cores. Finally, a description of the author's method of performing routine permeability measurements on laboratory samples, with a discussion of its validity, is given as a contribution to the control of the large proportion of the national output of magnetic material, usually neglected, which finds its way into small motors.

(4.1) A Tester for Whole Sheets of Transformer Steel

Although for many purposes laboratory tests are sufficient, they are inadequate to maintain a precise control of the quality of silicon-iron sheets used in transformers; because they are destructive, they can be applied only to a small proportion of the material. By the nature of hot-rolling processes there is always a greater variability in the magnetic quality of these sheets than the transformer designer is able to accommodate. It was to meet this state of affairs that a whole-sheet loss-testing apparatus has been designed. Models have now been in regular use for several years.

Previous testers which would accept entire sheets of magnetic steel, such as those due to Richter,[7,8] Schenck[9] and Sankey, differed from ordinary laboratory testers only in size and complexity. There is no new principle involved in the extension of standard methods of test to sheets measuring 8 ft × 3 ft if laboratory methods are permissible. In the present instance they were not. In particular, the rate of testing had to be far too rapid to allow of the sheets being weighed. The essential problems to be met were, therefore, first to provide some compensation for variations in sheet thickness; secondly, to ensure uniformity of flux density over the portion of the sheet whose losses are being measured without resort to any sort of assembly process; and thirdly, to replace the sensitive instruments of the laboratory by robust panel-mounted types for workshop use, and the manual adjustments customary therein by automatic controls.

(4.1.1) Mechanical Construction.

An idea of the general construction and magnetic circuit can be obtained from Fig. 8. Sheets on a conveyor belt pass through

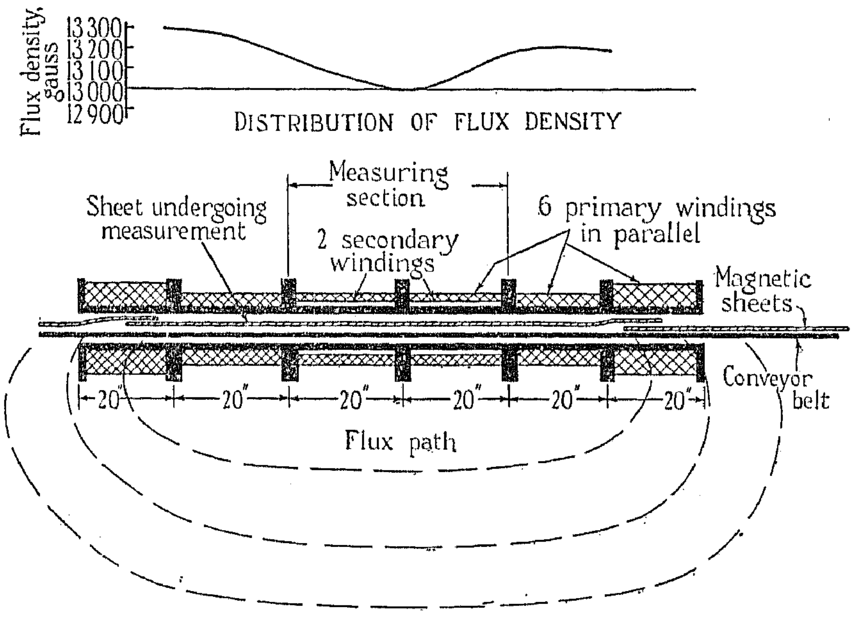


Fig. 8.—Whole-sheet tester.

a flat open-ended coil which has six sections in parallel. As in the Churcher or modified Epstein tester, therefore, the flux density in the two central sections—a length of 40 in—is remarkably uniform. The conductor section is graded, so as to give approximately uniform current density in all sections. Fig. 8 includes a typical flux-density distribution curve at $B = 13$ kG in the centre. Secondary coils to supply the wattmeter and voltage-control circuits are wound underneath the primary winding on the two central sections.

(4.1.2) Measurement of Loss.

Loss is measured in the portion of the sheet which lies within the two central sections of the coil, by supplying one element of a dynamometer wattmeter from the current in these two coils only (through a current transformer), and the other, through an amplifier, from one of the secondary coils. This is shown in

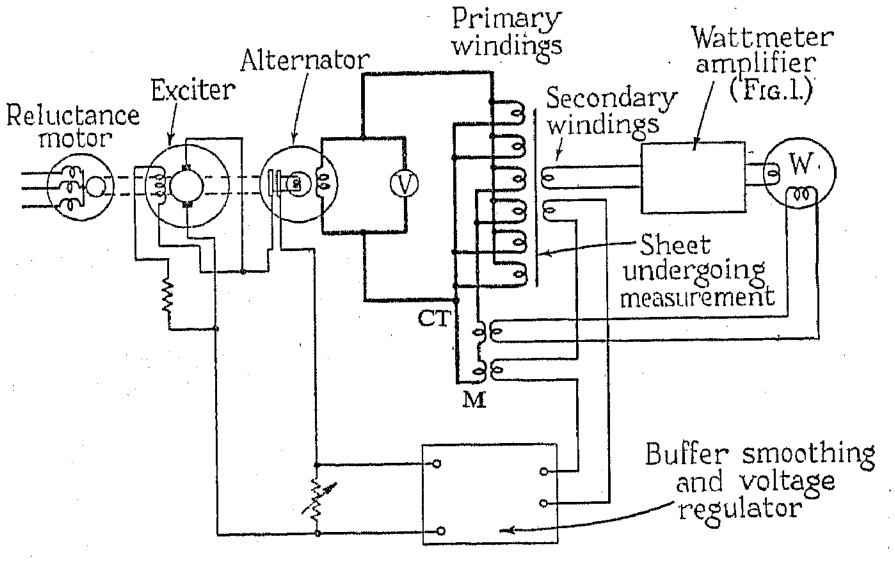


Fig. 9.—Whole-sheet tester—basic circuit.

Fig. 9. The amplifier, essential because the total loss is only a few watts, is that described in Section 2.1. Automatic protection, which short-circuits the current coils if these are overloaded, is

added: a high-speed relay is operated from the amplified voltage-drop across them. This is essential, for if a gap should inadvertently appear between sheets the current in the primary winding would rapidly rise to a value quite sufficient to burn out the current coil of the wattmeter. It is, however, usual to operate the tester with the conveyor moving continuously and a slight overlap between successive sheets.

This overlap has been shown in an intentionally exaggerated way in Fig. 8; the sheets are, of course, actually flat.

(4.1.3) **Compensation for Thickness Variation.**

Since our measurements can only be of total flux and loss, whereas we require to know—indeed, to set—flux density and to measure watts per pound, we must have some way of arriving at the cross-section of the sheet. Although the width is carefully controlled, the thickness must be expected to vary between limits of at least $\pm 10\%$ of the nominal value. If, therefore, we were merely to set our secondary voltage to a value corresponding to the nominal cross-section for a given flux-density, our actual flux-density would vary between the same limits, and the indicated specific loss would be in error by much the same amount; for a variation of $x\%$ in section would result in a variation of $-x\%$ in $\hat{B}$, approximately $-2x\%$ in total loss (which varies nearly as the second power of $\hat{B}$), and $(-2x + x) = -x\%$ in indicated watts per pound.

The method of compensation for this effect, while approximate, works well. If we denote by $\hat{B}$, $\hat{H}$, A_i and A_a the values of maximum flux-density and magnetizing force, cross-section of sheet and of coil air-space respectively for a sheet of nominal thickness, then the mean induced secondary voltage is

$$V_2 = n_2 f(A_i\hat{B} + A_a\hat{H}) \times 10^{-8}$$

where n_2 and f are the secondary turns and frequency. If V_2 is maintained constant while A_i, B and H vary, then

$$(\hat{B}\Delta A_i - A_i\Delta\hat{B}) - A_a\Delta\hat{H} = 0$$

i.e. the change in flux in the sheet is equal and opposite to that in the air-space.

Now if
$$\frac{\Delta\hat{B}}{\hat{B}} = \frac{1}{2}\frac{\Delta\hat{A}_i}{\hat{A}_i} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

the indication of the wattmeter would be invariable with thickness over the range for which the approximation $(\hat{B} + \Delta\hat{B})^2$ $\hat{B}^2 + 2\Delta\hat{B}$ is permissible; for the total loss is approximately proportional to $A\hat{B}^2$. This condition is achieved when

$$\frac{A_a}{A_i} = \frac{\Delta\hat{B}}{\Delta\hat{H}}$$

as will be seen by substituting in eqn. (5). For ordinary transformer steel this has a value of about 500 at $B = 13$ kG. To arrive at a condition where the wattmeter reading is not affected by variations in sheet thickness, it is therefore only necessary to provide sufficient air-space around the sheet and then maintain the secondary induced e.m.f. constant.

While it is convenient to be able to allow the coil to include a substantial area of non-magnetic material, it is also as well to provide for some adjustment. The mean secondary voltage of a mutual inductance whose primary winding carries the current of the central, or measuring, sections of the tester, is also proportional to $\hat{H}$. The tester coil is therefore made smaller in area than is necessary for complete compensation, and a voltage from a variable mutual inductance added in addition.

In practice, the compensation is set by trial and error, the value of mutual inductance being adjusted until the wattmeter reading is independent of sheet section over a wide range. For this purpose it is best to take a thick sheet, and shear portions from its width, adding these into the coil one at a time and checking the wattmeter reading. Table 1 gives a typical relationship between total loss reading and sheet section.

Table 1

WHOLE-SHEET TESTER: EFFECT OF VARIATION OF SHEET SECTION

Cross section of sheet, % ..	80	85	90	95	100	105	110	115	120
Wattmeter indication, % ..	92·6	95·5	98·9	100	100	99·6	99·4	97·8	97·0

Since the thickness variation of transformer sheets is limited by specification to $\pm 7\frac{1}{2}\%$, it will be seen that the compensation is adequate.

(4.1.4) **Compensating Circuit.**

The sum of the voltages from the secondary coil of the tester and of the mutual inductance is applied through a buffer amplifier to a rectifier and smoothing circuit, so that a direct voltage is produced proportional to the maximum value of a mixture of the flux and magnetizing-force. This is compared with a preset direct voltage, the difference automatically controlling the alternator field in a conventional way, using hard valves in the interests of good waveform. The circuit details, given in Figs. 9 and 10, require comment only in two respects. In the first place, the production of a direct voltage proportional to the mean of an alternating wave is by no means easy. The author could find no published reference to the problem, which requires for its solution that the rectifier shall be loaded up to a certain minimum value—hence the buffer amplifier and the 10-kilohm loading resistor following the rectifier in Fig. 10.

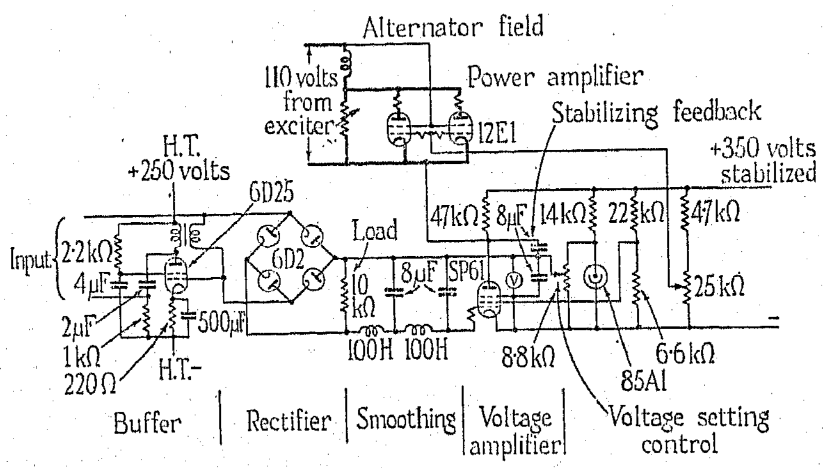


Fig. 10.—Whole-sheet tester—automatic voltage regulator.

Secondly, while it is quite easy to suppress instability in the regulator, it is a wise plan to allow a very small amount to remain, for a slight oscillation of the wattmeter pointer is an excellent antidote to stiction.

(4.1.5) **Performance of Tester.**

This tester has been operated continuously at speeds up to 10 sheets per minute. The limitation is then the speed of response of the dynamometer wattmeter. The author knows of no superior method of measuring a.c. power when the waveform is distorted; any such developments would be of considerable value in many other applications besides this.

As to accuracy, it must be remembered that this tester is not an absolute standard. It is calibrated by testing a number of sheets from which test pieces must subsequently be cut. In general, its readings are consistent within $\pm 3\%$; magnetic sheet is not usually sufficiently homogeneous to justify the assignment of a value of loss with a much greater precision. A variation in permeability will introduce an error proportional to that variation multiplied by

$$\frac{1}{2}\frac{\Delta\hat{B}}{\Delta\hat{H}}\frac{\hat{H}}{\hat{B}}$$

(taking, as before, values from the nominal curve at the nominal density). For typical transformer sheet, a variation of 20% in permeability is necessary to introduce a 1% error on the wattmeter. The equipment is practically self-compensating with respect to frequency variations.

(4.2) B/H Loop Tester

This device, of general utility, is of particular importance for general a.c. magnetic measurements and for routine tests on nickel-iron cores.

One of the earliest methods of arriving at magnetic properties was the use of a rectifying commutator. It suffered from the usual disadvantages of commutators, and the Ferrometer of Dr. Thal—still too little known in this country—was a considerable advance. This replaced the commutator by synchronous relays of very special design; this fact alone has placed it beyond the reach of the general public. Koppelman's cam-operated contact-making system[10] covers very similar ground, with certain omissions and extensions. In fact, however, it is perfectly easy to achieve equal accuracy with even greater flexibility by the use of a Carpenter relay if its driving circuit is designed with care. The apparatus to be described can easily be constructed from standard components at fairly small cost.

(4.2.1) **Principle of Measurement.**

Referring to Fig. 11, the flux in a toroidal sample is measured from the voltage induced in a search-coil of n_2 turns round the specimen, and the magnetizing force deduced from the value of the primary current, which is proportional to H if the specimen

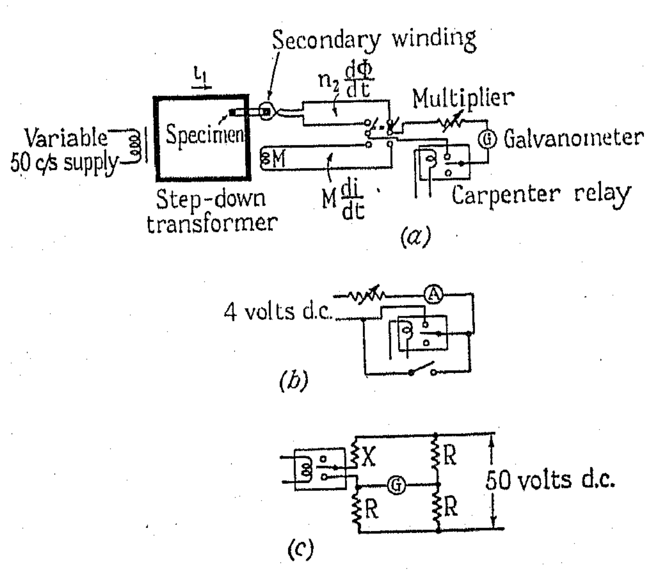


Fig. 11.—B/H loop tester.

is uniformly wound. The mutual inductance M has its primary winding in series with that of the specimen, so that the voltage appearing in its secondary winding is Mdi/dt. The measuring system is a plain galvanometer, or sensitive microammeter, arranged as a multi-range secondary voltmeter with one contact of a Carpenter relay in series, allowing current to pass for a portion only of each cycle. If Φ_1 and i_1 are the values of flux and current at the moment of "make," Φ_2, i_2 at the moment of "break," the mean galvanometer voltage will be

$2fn_2(\Phi_1 - \Phi_2)$ when connected to the specimen secondary winding

and $2fM(i_1 - i_2)$ when connected to the secondary of the mutual inductance.

It is now necessary only to be able to control the period of contact (referred to as the "dwell") and the moment in the cycle when it occurs (referred to as the "phase") to find out everything about the dynamic hysteresis loop with considerable accuracy. It is essential, however, that, once set, the dwell and phase should remain invariable for long enough for a measurement of Φ and i to be made, and also that dwell shall be measurable with accuracy.

(4.2.2) **Circuit of Tester.**

Fig. 12 shows how this is achieved. The coil of the relay is driven from a simple flip-flop oscillator. The mark-space ratio

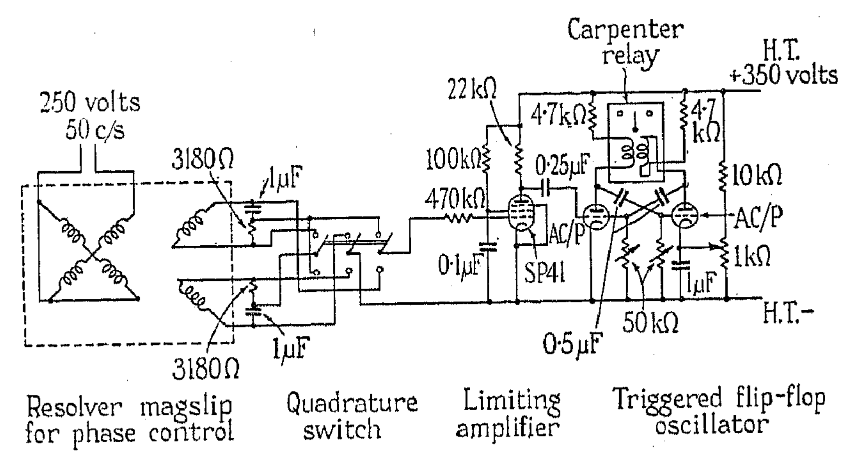


Fig. 12.—B/H loop tester—relay control circuit.

of the oscillator is adjusted, by the resistor shown, in the conventional way: this constitutes the dwell control, and enables the dwell to be readily varied from about one-fiftieth to just over one-half of a cycle. Ample stability is ensured by the use of a stabilized power supply. Phase control may be achieved by the use of a resolver magslip, in a circuit due to Patchett,[11] with an interposed limiting amplifier to make sure that the oscillator is triggered at exactly the same instant in each cycle. This feature is the only real precaution necessary to convert the Carpenter relay into a precision component. A simple circuit-switching operation changes the phase by exactly 90 electrical degrees.

Dwell is measured in two ways. For the first, the relay contacts are simply switched in series with a d.c. milliammeter, control resistor and battery. The resistor can be adjusted to give full-scale deflection with the relay contacts short-circuited, whereupon the deflection will be proportional to dwell when the short-circuit is removed. This method is shown in Fig. 11(b); in the actual instrument the various methods are selected by switching. The second method [Fig. 11(c)] provides for an alternative measurement of dwell with great accuracy at the one value of greatest general use: 50%, or one half-cycle. In this, the relay contacts are inserted in one arm of a Wheatstone

bridge which uses the galvanometer as balance detector. For 50% dwell, balance is obtained when the resistance X is

$$X = \frac{R^2}{3R + 2G}$$

This value, and the other bridge arms, are preset and built into the apparatus.

For convenience, a different set of multipliers is used for B and H measurement, while the change from one to the other, and the various dwell checks, are selected by switching. Any form of sample can be used, but the simplest is a packet of ring punchings, or a coil of strip. A single secondary turn gives ample sensitivity, while a single primary turn supplied from an inverted current transformer is a convenient form of supply. A toroidal mutual inductance is best for avoidance of interference, but gapped-core inductances are sometimes convenient.

(4.2.3) **Applications of B/H Loop Tester.**

In describing a few of the measurements which can be made with the B/H loop tester, it is convenient to use the symbol B for the instantaneous mean flux-density, and H for the surface magnetizing force, since these are the only measurable quantities. The dynamic hysteresis loop, upon which all measurements are made, differs from the static loop because, owing to the presence of eddy currents, the actual values of flux density and magnetizing force in the specimen vary from point to point, and with time.

$\hat{B}/\hat{H}$ Curve (a.c. induction).—This is obtained by setting the dwell to 50% and reading corresponding values of B and H with B a maximum, for varying induction.

Remanence.—With the dwell set to 50%, remanence may be measured by adjusting the phase control to obtain a zero reading of H and reading the corresponding value of B.

Apparent Coercivity.—The value of H for $B = 0$ is obtained by phase adjustment with 50% dwell, in a similar way.

Slope of B/H Curve.—With the dwell set to a small value, corresponding values of ΔB and ΔH can be read at all points on the loop. The slope is a maximum when ΔB is greatest for a given dwell.

Routine Testing of Current Transformer Cores.—The apparatus has the advantage over bridge methods that single-turn primary and secondary windings will usually suffice. It measures $\hat{B}$ instead of the fundamental component, but at low flux-densities the difference is not usually significant. The value of $\hat{B}$ is set with 50% dwell and appropriate phase control, whereupon the in-phase and quadrature values of H are immediately obtained. It is best, with B unchanged, to adjust the phase control until a zero reading of B is obtained. The corresponding reading of H is the quadrature component, while operation of the quadrature switch gives the in-phase component.

(4.3) Routine Permeability Measurements

The permeability of magnetic sheets is a quantity of importance, in particular, to the designer of induction motors. For its measurement, numerous types of permeameter have been developed; they all operate on a reversal of magnetizing force, and all measure the flux density by means of a ballistic galvanometer connected to a search coil. Since it is impossible for any of them to be used with a predetermined value of B—although the Iliovici types can work with a predetermined value of H—routine testing for permeability is far more tedious than for loss. It is, in fact, rarely performed.

For many years it has been known that, in suitable circumstances, the values of permeability deduced from the peak flux and peak current measured on a sample at 50 c/s differ little from those obtained on d.c. permeameters. The difficulty of extending this work has lain chiefly in determining peak currents, without the use of a rotating commutator. While the method described in the previous Section is suitable, an even more convenient arrangement is the use of an amplifier-operated peak-reading ammeter in conjunction with a Lloyd–Fisher square.

(4.3.1) **Arrangement of Circuit.**

Fig. 13 shows the complete circuit which the author finds convenient. The square is of conventional construction, and is

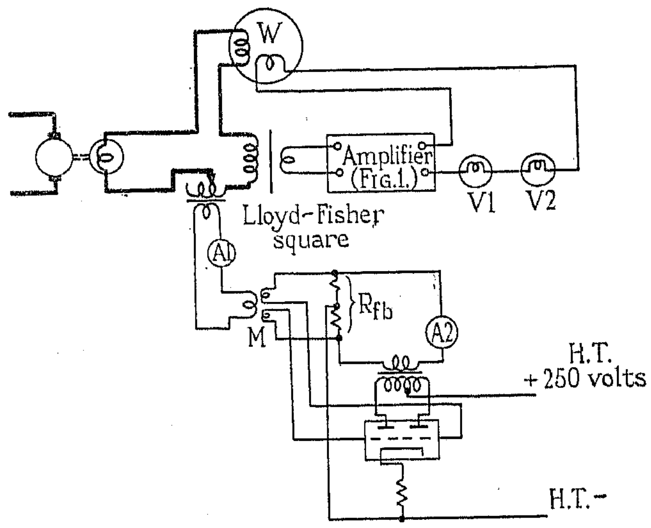


Fig. 13.—Permeability measurement with the Lloyd–Fisher square.

supplied from an alternator of low impedance, driven by an electronically controlled d.c. motor which provides a frequency range of 0–100 c/s. Frequency is measured on a Maxwell bridge. The wattmeter amplifier described in Section 1 is used to supply not only the voltage coil of the wattmeter but also a rectifier and a thermal milliammeter—both scaled in volts—the one to read peak flux and the other to monitor the form factor. For the measurement of primary current a variable-ratio current transformer supplies both a moving-iron instrument A_1, to read r.m.s. values, and the primary of a mutual inductance. The amplifier voltmeter, shown in the circuit, is of simple design, and provides ample negative feedback: its output milliammeter, A_2, reads peak primary current, according to the principles expounded in Section 4.2.

(4.3.2) **Method of Operation.**

The value of $\hat{B}$ is set from the secondary mean voltage and the power measured in the ordinary way. In addition, a reading of peak primary current is recorded. From this, $\hat{H}$ is readily calculated. It can, incidentally, be employed in making a correction for air flux, or even this can be avoided by the conventional use of another mutual inductance. Apart from the fact that no more work is required than for a normal loss-test, this method has the considerable advantage that measurements are made at a preset value of $\hat{B}$.

(4.3.3) **Comparison with "D.C." Permeability.**

To justify the use of this highly convenient method it is necessary first to establish its validity on a symmetrical sample, and secondly to make sure that the construction of a Lloyd–Fisher square does not introduce further errors.

Absolute Measurements.—A comparison between the $\hat{B}/\hat{H}$ curve measured on ring samples by conventional methods and deduced from 50-c/s readings has been made before.[12] The author repeated the work, using the apparatus just described for the a.c. measurements, on two samples, one of 0·2% and the other of 1·5% silicon steel, both 0·016 in thick. The results are given in Table 2, the "d.c." values being obtained by reversals in the usual way.

Table 2

COMPARISON OF "A.C." AND "D.C." VALUES OF B FOR A COMMON H, ON TWO RING SPECIMENS

H	Specimen No. 1 1·5% silicon		H	Specimen No. 2 0·2% silicon	
	B_{dc}	B_{ac}		B_{dc}	B_{ac}
0·9	7 000	5 500	0·9	6 500	—
1·0	8 000	6 400	1·0	7 500	3000
1·2	8 900	7 800	1·2	8 900	4 500
1·6	10 100	9 800	1·6	10 300	7 100
2·0	11 000	10 900	2·0	11 200	9 500
3·0	12 250	12 400	4·0	13 400	13 300
4·0	13 000	13 100	6·0	14 300	14 220
8·0	14 100	14 100	10·0	15 300	15 250
20·0	15 100	15 100	16·0	15 860	15 870

Above $B = 12$ kG, the agreement is within the limits of experimental error—indeed, the results reflect favourably on the accuracy of the amplifier systems employed. Below $B = 12$ kG, the a.c. values fall off sharply.

Measurement on Lloyd–Fisher Square.—Accordingly, an extensive series of tests was carried out on samples assembled in a Lloyd–Fisher square. The results of these, given in Table 3,

Table 3

COMPARISON OF A.C. PERMEABILITY MEASUREMENT IN A LLOYD–FISHER SQUARE WITH D.C. MEASUREMENTS

Silicon content	Thickness	No. of tests	Difference		
			$\hat{B} = 10\,000$	$\hat{B} = 13\,000$	$\hat{B} = 15\,000$
%	in		%	%	%
0·2	0·016	1	1·3	0·7	0·0
0·2	0·025	2	12·0	4·4	1·7
0·75	0·016	1	2·5	0·7	−0·6
0·75	0·025	1	8·0	2·7	0·3
1·5	0·016	6	3·4	0·9	0·6
1·5	0·025	5	8·0	2·2	0·9
3·8	0·016	9	6·2	1·5	0·3
4·2	0·014	4	3·2	0·7	−0·2
4·2	0·018	2	11·1	2·3	−0·2

The Table gives the percent difference between $\hat{B}$ as measured on alternating current in a Lloyd–Fisher square and d.c. values measured in an Iliovici permeameter for a common value of $\hat{H}$. It is positive when the d.c. value exceeds the a.c. value. Each figure is the average of a number of separate tests, as shown.

where each figure is the mean of measurements on several different samples, bear out closely the findings of the previous Section. Permeability measurements at low densities are of no real interest to the machine designer, who finds most of his magnetizing current accounted for by an air-gap of unity permeability and those sections of his magnetic circuit which may be saturated. It is therefore deduced that over the range of flux densities where measurements are normally required for power engineering work the a.c. measurement of permeability is fully justified. It is difficult, indeed, to argue that even where there is a discrepancy the d.c. values are of greater interest, for the conditions of test by reversals bear no resemblance to the magnetic state of any a.c. machine or transformer. The author would therefore urge the establishment of a.c. permeability measurement as a standard method of control in any future British Standard on this subject.

(4.3.4) **Theory of A.C. Measurements of Permeability.**

When permeability measurements are made by the method of reversals, the flux distribution across the section of the sample may be initially far from uniform, but the ballistic galvanometer reading is always taken after ample time has been allowed for ultimate uniformity to be attained. If the value of $\hat{B}$ for the same $\hat{H}$ when the sample is excited on alternating current is less than by the method of reversals—it obviously cannot be greater—the reason can only be that the flux density has not achieved uniformity when Φ is a maximum. An examination of Table 3 shows clearly that the effect, increasing with the thickness of the specimen, must be due to internal eddy-currents.

That this is indeed so at low values of B, where the permeability is high, cannot be doubted; direct experimental evidence has been produced by Brailsford[13] in a remarkable paper. Brailsford also gives the classical curve showing the dependence of $\hat{B}$ on permeability. When it is remembered that during every hysteresis cycle the slope of the loop for nearly all common types of magnetic steel reaches values of the order of 30 000, the results in Tables 2 and 3 are easily accounted for. The condition at higher densities resembles more closely that obtaining in the method of reversals, for the very low differential permeability permits the rapid equalization of flux distribution across the section of the sample before the flux reaches its maximum.

(5) DIELECTRIC MEASUREMENTS

While there is increasing appreciation of the value of measurements of power factor in assessing the quality of the insulation of the windings of high-voltage machines and transformers, it cannot be pretended that a.c. bridges are ideal methods for factory use. In order to arrive at the power factor it is necessary to measure in addition an unwanted quantity—the capacitance. The instrument to be described has the advantage that it measures power factor or loss angle directly; it can also be used at almost any voltage.

(5.1) Principle of Power-Factor Tester

The Carpenter relay is again the basis of measurement, this time being used as a polarized rectifier. Referring to Fig. 14,

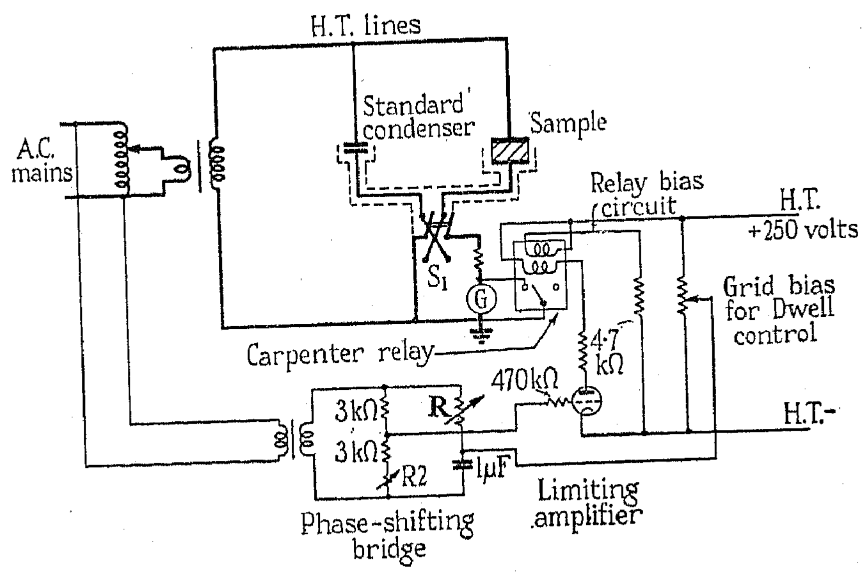


Fig. 14.—Power-factor tester for insulating materials.

it can be seen that the current from the specimen is passed through a d.c. galvanometer, shunted by the contacts of the polarized relay. The relay coil is supplied from a limiting amplifier, which in turn is excited through a phase-shifting circuit from the mains. With this precaution the relay performance will be found to give quite adequate consistency.

If the period of contact of the relay extends equally on both sides of a current zero the mean galvanometer current is zero, but if this is not so the galvanometer will read either positive or negative. In order to bring about a null reading, therefore, irrespective of the magnitude of the current, it is necessary only to adjust the variable resistance R in the phase-shifting circuit; this can be graduated in terms of tan δ, where δ is the loss angle, or complement of the phase angle, of the sample. In order to establish the point $\tan\delta = 0$, from which the scale may be determined by calculation, the sample is replaced by a standard condenser whose factor is known. It is, in fact, a convenience to provide a switch S, to make this change periodically; trimming adjustments for zero are then made by small variations in the resistance of R_2.

(5.2) Precautions in Operation

Several precautions are necessary in the use of the apparatus.

(5.2.1) Screening.

As with all high-voltage measurements, it is essential to make certain that the only current which passes the measuring system—in this case the galvanometer—is that which originates at the low-voltage electrode of the sample. All the other currents must be lead away to earth by a comprehensive screening system, which is indicated in Fig. 14. Particular care must be taken of the standard condenser. For voltages up to about 10 kV a vacuum condenser is ideal, but it is essential to conduct away, by means of a conducting clamp, the leakage current over the glass envelope which would otherwise shunt the capacitance.

There will be a potential difference between the screen and the low-voltage electrode of the sample, or of the standard condenser, but this is far too small to give rise to the errors which have to be guarded against, by Wagner earthing devices, in a.c. bridge work.

(5.2.2) Phase Angle of Supply Transformer.

The supply transformer forms part of the measuring system, and if the load imposed by the sample is an appreciable fraction of its rating—which is as much as to say, if its leakage-reactance potential drops are appreciable—it is necessary to standardize with the sample in circuit, but with its low-voltage electrode earthed: the switch S_1 in Fig. 14 performs this function. An alternative which is quite feasible is to energize the phase-shifting circuit direct from the h.v. line, or through a potential transformer.

(5.2.3) Waveform Errors.

This system measures the mean current, whereas bridge methods take only the fundamental current into consideration. It is therefore to be expected that there will be slight discrepancies between the two methods if the sample has non-linear characteristics. All insulating materials are non-linear to some extent, but no substantial errors have been found on this score; on the other hand, care does have to be taken that the supply voltage does not depart appreciably from a sine wave, since it is being differentiated.

(6) ACKNOWLEDGMENTS

The author wishes to record his sincere gratitude to his colleagues, in particular to Mr. B. Atkins and Mr. E. A. Hall, whose enthusiastic co-operation has made possible the development of the various types of apparatus described; also to Mr. H. S. Holbrook, to whom many of the concepts underlying the section on magnetic measurements are due; and to Mr. H. Dreghorn, and the Directors of the British Thomson-Houston Co., Ltd., in particular Mr. H. L. Satchell and Mr. H. Jack, for continual encouragement and help.

(7) REFERENCES

(1) CASSON, W., and LAST, F. H.: "Ultra-High-Speed Relays in the Fields of Protection and Measurement," *Journal I.E.E.*, 1947, **94**, Part II, p. 667.

(2) ARNOLD, A. H. M.: "Performance Limits in Electrical Instruments," *Proceedings I.E.E.*, Paper No. 1128 M, March, 1951 (**98**, Part II, p. 701).

(3) MACFADYEN, K. A., and HILL, N. D.: "A Precision Thermionic A.C. Test Set," *G.E.C. Journal*, 1941, **11**, p. 182.

(4) MAXWELL, J. C.: "A Treatise on Electricity and Magnetism," (Clarendon Press, Oxford, 1904, Third edition), Vol. 2. Arts. 775 and 776.

(5) CASE, N. P.: "A Precise and Rapid Method of Measuring Frequencies from Five to Two Hundred Cycles per Second," *Proceedings of the Institute of Radio Engineers*, 1930, **18**, p. 1586.

(6) THOMSON, J. J., and SEARLE, G. F. C.: "A Determination of '*v*', the Ratio of the Electromagnetic Unit of Electricity to the Electrostatic Unit," *Philosophical Transactions of the Royal Society of London*, A, 1890, **181**, p. 583.

(7) RICHTER, R.: "Vorschlag zu einem neuen Eisenprüfapparat," *Elektrotechnische Zeitschrift*, 1902, **23**, p. 491.

(8) RICHTER, R.: "Eisenprüfapparat für ganze Blechtafeln," *ibid.*, 1903, **24**, p. 341.

(9) SHENCK, W. E.: "The Absolute Measurement of the Energy Loss in a Single Strip of Sheet Iron or Steel under Alternating Magnetization," *Proceedings of the American Society for Testing Materials*, 1942, **42**, p. 595.

(10) KOPPELMANN, F.: "Verfahren zur Messung annähernd rechteckiger mit grosser Geschwindigkeit durchlaufener Magnetisierungsschleifen von Ringkernen," *Archiv für Elektrotechnik*, 1944, **38**, p. 141.

(11) PATCHETT, G. N.: "A Simple Phase Shifting Device," *Electronic Engineering*, 1951, **23**, p. 310.

(12) WEBB, C. E., and FORD, L. H.: "Alternating-Current Permeability and the Bridge Method of Testing," *Journal I.E.E.*, 1935, **76**, p. 185.

(13) BRAILSFORD, F.: "Investigation of the Eddy-Current Anomaly in Electrical Sheet Steels," *ibid.*, 1948, **95**, Part II, p. 38.

DISCUSSION BEFORE THE MEASUREMENTS SECTION, 1ST FEBRUARY, 1955

Prof. J. Greig: The part of the paper dealing with the application of negative-feedback amplifiers in magnetic measurements is of particular interest to me. My first contact with such amplifiers was also in connection with a magnetic-measurement problem, but not wattmeter measurement. It was the measurement of peak flux, the problem being to obtain the accurate rectified mean value of an induced voltage where that voltage was so small as not to be susceptible of conventional methods of rectification. At that time we built a negative-feedback amplifier which had good enough phase and amplitude response to give correct reproduction of the waveform for the purpose of rectification. I have therefore been particularly interested to read the description of the author's apparatus and to appreciate that magnetic-loss measurements can now be made with such high accuracy utilizing the negative-feedback-amplifier technique in association with a wattmeter. I should like to ask the author whether he has made use of the interesting negative-feedback-amplifier technique, which, so far as I know, was developed by one of the sheet-steel making organizations, and in which the negative-feedback principle is used to maintain waveform in magnetic measurements. The principle involved is that of using the test specimen as the output transformer of the negative-feedback amplifier, thereby using the feedback to compensate for the distortion of the test specimen as well as to compensate for inherent distortions in the amplifier itself.

The use of the polarized relay for purposes of frequency or speed measurement is also interesting, and I should like to ask the author whether he can say anything about the transient response of such a speed-measuring device; whether, in fact, it is possible to measure accelerations and decelerations with it.

Dr. A. H. M. Arnold: I am interested in the author's amplifier-wattmeter but am not clear whether this instrument is intended for use at power frequencies only or for use also at audio frequencies. The author's remarks in Section 2.2 appear to suggest that he contemplates use at audio frequencies up to 5 000 c/s and above, but the circuit shown in Fig. 3 would not be suitable for a wide range of frequencies on account of the errors of the current transformer. It may be remembered that the principal limitation in the performance of the original amplifier-wattmeter designed by MacFadyen and Hill arose from the transformer error. This limitation may be avoided by the use of two amplifiers, one for the moving coil and one for the fixed coils of the wattmeter, and a further advantage may be gained by using only one indicating instrument for voltage, current and power measurements. If the instrument is intended for use at audio frequencies the quality of the voltage divider becomes of importance, and it may be difficult to get the requisite performance with a resistance divider having the value of 10 kilohms per volt suggested by the author. If a capacitance divider is used the current consumption becomes large at the higher frequencies. In our latest model amplifier-wattmeter for use at frequencies up to 20 kc/s, the voltage divider has a resistance of only 1 kilohm per volt, and even then has required resistance units of special construction having very small time-constants.

The accuracy of the circuit shown in Fig. 4, for low-power-factor measurements, depends on the purity of the mutual inductor used. Some care would be necessary to obtain adequate secondary voltage from the mutual inductor without the introduction of a phase defect on account of capacitance in the windings. It would be interesting if the author would give dimensional details of the instrument used.

Prof. F. Brailsford: The development of the whole-sheet testing equipment is a considerable achievement. The tester calls to mind the standard type for strip specimens due to Churcher and also the Dannatt single-strip tester, since all three employ sectionalized primary windings with the object of improving the uniformity of flux density in the specimen. The author is, however, faced with the difficulty that the only return circuit for the magnetic flux in the sheet under test is through air-leakage paths. He evidently endeavours to reduce the reluctance of these paths by overlapping the preceding and the following sheets. In the circumstances the author's curve of flux distribution, given in Fig. 8 as a typical example, is surprisingly good, since it shows, for a thin sheet, which may be 3 ft wide and 8 ft long, a variation in flux density of only about $\pm 1\%$ of the mean value. It is also surprising that the flux density is lowest at the centre and increases towards the ends, for the opposite effect might be expected with an open-ended solenoid, even with sectionalized windings. This result might have been achieved by reducing the number of turns on the end sections of the primary winding. Again some variation in the flux-distribution curve might be expected in successive sheets, since considerable variability in the overlapping contacts of sheets, which are never quite flat, must occur.

This tester exists because, in the past, experience has shown substantial variations in magnetic quality in the individual sheets in a given batch and its function is to sort them. This is, in fact, done by comparing the central sections of successive sheets, since the measurements are made only on a length of about 40 in of each sheet. This might not be sufficient if the individual sheets were magnetically inhomogeneous, particularly if they suffered from erratic variations in quality between the centre and the ends. The author has no doubt satisfied himself on this point.

Mr. F. S. Edwards: Section 5 describes one of the many applications of the Carpenter relay—in this instance for the measurement of power factor. Although the scheme is ingenious, it seems to me to be rather elaborate, and with the elaboration goes the risk of error. However, the final proof of a device of this kind is the experience of the user, and the author has evidently satisfied himself of its convenience and reliability. The introduction to this Section contains two statements on which I venture to differ from the author. First, although a.c. bridges may not be ideal for factory use they are extremely simple in construction and operation, and for most purposes their errors are either negligible or are readily computed. I have not found any difficulty in using them or in teaching others to do so. Secondly, the author's suggestion that the capacitance is an unwanted quantity is at variance with my own experience. There are many occasions where the power factor by itself gives an incomplete—or possibly even a misleading—answer, and the permittivity is frequently of as much interest as the power factor; indeed it sometimes happens that the power factor is the unwanted quantity and that it is only the capacitance that is required.

Mr. C. Ryder: When I glanced through the first pages of the paper I felt convinced that I had come across the tester's dream: a wattmeter with a large range both in voltage and in current, full-scale values such as $0 \cdot 3$ watt to 30 kW, and a high-impedance voltage coil. On reading closer, however, I must confess that my enthusiasm was damped a little by realizing that the virtue of the instrument really lies in the amplifier.

The crux of the matter seems to me to be not so much whether an amplifier can be designed and made to do the job—that goes without question—but rather whether, when an amplifier is coupled with such a universal type of instrument as a wattmeter and put in the hands of a test department for use by people who are other than electronic enthusiasts, it can be looked upon with the same degree of reliability as you would

look upon an ordinary standard voltmeter or wattmeter in normal laboratory use.

In the circuit shown in Fig. 2 the author describes a method of testing the wattmeter in conjunction with resistance and mutual inductance, but if I read it correctly the test only comprises a check at unity and zero power factor. Does this mean that tests at other power factors are unnecessary?

I should also like to know whether he has obtained an N.P.L. certificate for his universal test set (Fig. 3).

Regarding Fig. 4, I should like to ask what happens with regard to small changes in circuit values of voltage, current, or power factor, because from the diagram it seems that they may bring about a relatively large swing in the phase position of V_1.

Also, what is the method adopted in determining the most convenient adjustment of the voltage to put V_1 in the right phase position, and is any attention by the tester necessary to hold it?

Mr. R. G. Martindale: I should like to ask the author whether the low-power-factor circuit shown in Fig. 4 has been used for iron-loss testing where the presence of higher harmonics might lead to trouble owing to impurity in the mutual inductance. The introduction of a mutual inductance removes the problem of phase-angle error from the current transformer to the inductance itself, and in the case of iron-loss testing, where higher harmonics may be present, capacitance between turns might lead to an error under certain conditions.

As to the comparison of permeability measurements under a.c. and d.c. conditions of test given in Table 2, there is a difference of about 30% for a 0·2% silicon sheet at a d.c. flux density in the region of 10kG. On the other hand a similar sample given in Table 3 shows a difference of only 1·3% at the same flux density. I should have expected rather the opposite state of affairs, namely that there would have been a bigger difference in the case of the Lloyd–Fisher test, where there are air-gaps in the magnetic circuit, than with ring tests, where the position is not obscured by air-gaps. No reference is made in the paper to the latter point.

At a density of about 13kG the permeability of these materials is about 3 000, and an error of 10% in the effective value of H in the material, arising from the presence of air-gaps, would lead to an error in B of about 3%. Now a 10% error in H with a 100cm Lloyd–Fisher square is equivalent to 10cm of iron, which at a permeability of 3 000 is the same as only a little over 1 mil of air distributed over eight joints. Even allowing for the corner-piece overlaps, this is not very much. I would agree with the author's remarks on the results of the Lloyd–Fisher tests given in Table 3 that the differences are due apparently to internal eddy currents, but I think the result is rather surprising in view of the presence of joints. I feel that the agreement might perhaps not be so good with materials of higher permeability. Whilst the skin effect in the test sample increases with the permeability, as pointed out by the author, the proportion of the applied m.m.f. which is expended on the joints also increases, and errors due to the latter cause might also be expected. Has the author made any measurements on high-permeability materials to confirm the existence of two separate effects?

I think a number of people may have been impressed by the extensive use by the author of the Carpenter polarized relay, and I should like him to enlighten those of us who have been stimulated by this work by giving us some further information as to its reliability in service. Points which come to mind are contact rebound, susceptibility to mechanical vibration (which might be quite important in respect of equipment used in the factory), length of operational life, maximum current rating and the order of the transit time.

Mr. F. H. E. Myers: I should like to know the intrinsic accuracy of a commercial type of instrument based on the Carpenter relay. I understand that there is an instrument of foreign origin which can be relied upon to an accuracy of 0·01%, and I wonder whether a frequency indicator based on the Carpenter relay would be comparable, and whether it is available in a commercial model.

The author gave an important place to electronics in the measurement of electrical quantities. Would he like to say something about the possibility of integrating electrical energy by using a dynamometer wattmeter and converting its deflection by electronic devices into an integrated quantity over a given period of time?

It was 100 years ago, in 1855, that Siemens gave us practically the first induction motor. Fifty years later, in 1905, the induction-pattern rotating instrument was brought into commercial use. This is 1955 and perhaps it is a portent of change. The electronic device or artifice I have in mind consists of an RC circuit, C being variable and driven by the instrument spindle. A stabilized d.c. supply could be applied and the condenser discharged at a preset voltage. The total pulses of charge and discharge could be counted electronically and would measure an integrated quantity.

NORTH-WESTERN MEASUREMENTS GROUP, AT MANCHESTER, 22ND FEBRUARY, 1955

Mr. D. A. Langford: I shall confine my remarks to the whole-sheet tester. From Table 1 it is seen that a variation in thickness of $\pm 7\frac{1}{2}$% gives only a 0·5% error in the total iron-loss for a sheet of a particular material, with the mutual inductance adjusted by trial and error to give the optimum thickness compensation. This thickness compensation requires that $\Delta B/\Delta H$ at 13kG is correct for the sheet being tested. Now $\Delta B/\Delta H$ for hot-rolled silicon steel at 13kG varies from 560 for a 0·4% silicon-steel to 140 for a 4% silicon-steel. Because of this variation the thickness compensation alone allows errors of about the ± 3% mentioned as the overall consistency of readings, unless the mutual inductance is changed for different grades of material.

Owing to the thickness compensation, the flux in the steel at 13kG is only 85% of the total flux linked with the magnetizing coil; thus there is a considerable departure from the sinusoidal flux which is desirable for iron-loss measurements.

Mr. F. W. Taylor: First, I am very pleased to see that the author has made such good use of the Carpenter relay. Although this instrument has been available for some time and should be well known, it is not nearly as popular as its usefulness should make it. Secondly, I agree with the author that we can no longer afford to neglect the advantages which electronic techniques give to measurement. Nowadays the electronic engineer can do almost anything by juggling with electronic components and circuit techniques, and it is so easy for him to diagnose a fault, replace a valve, or even remove other components with a soldering iron.

In the factory, however, expensive plant and material are often vitally dependent on some small electronic component, and it is not always possible to have an electronic technician standing by. Consequently a very high standard of reliability is required or, alternatively, simple checks coupled with adequate instructions to the works staff are essential if full use is to be made of the electronic techniques in factory measurement.

To illustrate this point I would mention that the organization with which I am associated has an electronic instrument measuring the hydrogen content of the inert atmosphere of an annealing furnace. During the night this instrument ceased to

function, and the furnace could not be shut down because it contained a very expensive charge of cold-rolled steel at 800°C. Further, a high hydrogen content would be detrimental technically and would increase the explosion hazard. The fault was simple—a thermionic valve had failed and was replaced very quickly by the engineer, but the works superintendent was in a state of considerable apprehension until he appeared.

We have found that an amplifier very similar to that shown in Fig. 1 is extremely useful in cases where a high impedance is required, such as routine iron-loss testing by the Lloyd–Fisher square. In our case provision was made in the instrument for calibrating the rectifier voltmeter and checking the amplifier itself.

Incidentally, in view of the difficulty of making accurate and stable mean-voltage voltmeters for the checking of peak flux-density, has the author tried the Carpenter relay and a moving-coil instrument for this purpose?

A transformer sheet grader was built some two or three years ago on the lines of that described in Section 4.1. In our case the loss in the sheets to be graded was compared with that of a standard sheet specially chosen to have average characteristics for that grade, i.e. thickness, permeability and loss. This standard sheet was held stationary inside coils identical with those of the grader, and a house service-meter with special windings was used as a timing relay. When a sheet entered the measuring coils a contact switched on the supply to both sets of coils so that the timing relay and the measuring watt-hour meter started to rotate. After a given interval, during which all but the ends of the sheet under test passed through the measuring coils, the timing-relay contacts closed, both sets of coils were de-energized and both meters stopped. The reading on the disc of the measuring meter, marked in grades, was thrown on to a screen for 3 sec. After this both meters returned to zero and the whole cycle was repeated.

It was noticed that without thickness compensation the grading was correct only to within about $\pm 7\%$. Thickness compensation was obtained by adjusting the ratio of the voltage on the grader coils to that on the timer coils in accordance with the thickness of the sheet being tested. It was then found that 60% of the sheets could be graded to within 2%, and the remainder, except for very occasional freak cases, to approximately $\pm 3\%$. Weighing the sheet to obtain the thickness, however, slowed down the process too much for a production line.

The thickness-compensation argument is ingenious, but, in view of the wide variations in permeability over the waveform and for different grades of steel, the author's results must be based on a compromise. In any case, if the mutual inductance used for adjustment is too large the argument fails.

Finally, in Section 4.1.3 the author states that "the thickness must be expected to vary between the limits of at least $\pm 10\%$ of the nominal value." With this I agree, but at the end of the Section he makes the claim that since the thickness is limited by specification to $\pm 7\frac{1}{2}\%$, Table 1 shows that compensation is adequate. In practice, where variations of $\pm 15\%$ are encountered, the wattmeter indication would only be within 5% for really thin sheets, which is equivalent to three-quarters of a grade.

Mr. R. G. Martindale: The method by which thickness variations between individual sheets are compensated for in the single-sheet tester, by adjusting the supply voltage automatically so that a constant voltage is induced in a B coil having a large air flux linked with it, is most ingenious. This correction is made on the assumption that the B/H characteristics of the materials tested are reasonably alike. It is my impression that single-sheet testing is, at present, confined to sheets having a loss of 0·92 watt/lb and lower loss at $B = 13$ kG and 50 c/s, and the figures of $\pm 3\%$ mentioned in the paper—i.e. plus or minus one-half grade—represent a big step forward in providing sheet material of more uniform quality for the electrical industry. It is perhaps not generally known that a large proportion of the transformer grades of sheet made in this country are individually sheet-tested on equipment made to the author's designs.

The extension of single-sheet testing to the slightly worse grades of sheet, up to a loss value of say 1·15 watts/lb at $B = 13$ kG and 50 c/s, should, however, result in a more uniform product becoming available for use in large rotating machines, and would constitute a further step forward. Has the author contemplated or carried through any such extension of single-sheet grading? Does he think that one setting of the thickness compensation would suffice for these additional grades also, or would an additional setting of the mutual inductance shown in Fig. 9 have to be provided?

I was particularly impressed by the B/H loop tester, which relies on the Carpenter polarized relay and its associated control circuits. Anyone who has had to make hysteresis-loop measurements, particularly on high-permeability materials with a rectangular form of loop, will appreciate what a useful device this is, and I think that there would be considerable scope for its use in general laboratory work as well as in works test areas. Regarding the lantern slide showing the rectangular hysteresis loop for a nickel-iron alloy, I was most interested in the re-entrant nature of the steep sides of the loop, but feel that it might be premature to attribute this to an inherent property of the material without further investigation. Re-entrant loops of a discontinuous nature have, however, been obtained in the United States with some nickel-iron alloys.

Anyone reading the paper cannot fail to have been impressed by the versatility of the polarized relay, and I think that its use, together with that of negative-feedback amplifiers, as shown by the author, points the way to many interesting future developments in measurement work generally.

Mr. E. Rawlinson: In the measurement of power at low power-factors the following facts must be realized: The use of an amplifier to drive the wattmeter does not increase the scale reading, and in order to measure with a power factor of, say, 0·05, the instrument must be sufficiently sensitive itself to give a reasonable deflection with full rated currents in the coils at this power factor. This means that a sensitive wattmeter must be used, and in this respect, therefore, the use of an amplifier does not mean a more robust instrument. This power factor of 0·05 occurs when iron losses in 4% silicon-transformer steel at a peak flux density of 15 kG are measured, and it is therefore in the range of routine measurement. An accuracy of 1% is only obtained at this value if the phase defect between the current through the wattmeter voltage coil and the induced voltage in the secondary winding is not greater than 0·02°. This means, in the amplifier of Fig. 1 which has a feedback factor of 60 dB, that the open-loop phase angle must not be greater than about 20°, and the capacitor C must be chosen so that this is true. It is not true, as stated in Section 2.1, that 60 dB of feedback necessarily means phase errors of less than 0·1%. The attenuator, when used, being outside the feedback loop, should also have a very small phase error.

Section 4.3.4 describes the effect of eddy currents or magnetic skin effect on the a.c. measurement of permeability, but in using a Lloyd–Fisher square for these measurements other effects are present and important. Two sources of error are the corner pieces with their air-gaps and the lack of uniformity of flux density along the sample. All these errors act in the same direction and produce readings of H higher than the true value, and all are greater for the higher-permeability alloys. Tests made recently on samples of the best cold-reduced directional transformer steel gave readings of H about three times the d.c.

value at $B = 10$kG, where the permeability is very high, and differences of only a few per cent at $B = 17$kG, where the permeability is much reduced. Measurements of a.c. permeability on these steels could therefore be carried out only after establishing uniform flux conditions—e.g. by sectionalizing the magnetizing coils—and even then the effects due to eddy currents would still be present.

Table 3 shows very good agreement for hot-rolled materials, especially at $B = 15$kG. The author uses an Iliovici permeameter for the d.c. measurements, and I wonder if he could give some idea of the accuracy of this method, since it is stated by Astbury* that errors up to 100% in H occur, and at $B = 15$kG this would mean an error of 10% in B.

Will the author describe the use of the mutual inductance when applied to the 2-wattmeter method of power measurement as mentioned at the end of Section 2.4? When, in a near-balanced system, one wattmeter has a low power-factor, the other will have a good one, and errors in the first wattmeter will have a small effect anyway, particularly if low-power-factor wattmeters are used.

Mr. J. E. M. Coombes: The author mentions that there has been some distrust of electronic techniques amongst measurements engineers, and that he has been at pains to ensure reliability. Had such precautions been more often observed in the past there is little doubt that acceptance of electronic devices outside laboratories would have been greatly accelerated, not only in the special field of measurements but in engineering generally.

Some idea of the sort of precautions necessary may be seen in Fig. 1. From the numerical data in the text it is easily shown that if the total gain of the valve stages should increase by 25%—the kind of change which may easily occur when valves are replaced—the overall gain of the circuit would nevertheless be increased by only about 0·02%. This desirable result is obtained, of course, by the liberal use of feedback. However, if the fraction of the output fed back is itself susceptible to change, then we have merely exchanged one source of variation for another. In Fig. 1, for example, if the fraction fed back should vary by 1% the overall gain would also change by about 1%. In the author's case, the output current I of the amplifier flows directly through a 50-ohm resistor, R_{fb}, and the quantity fed back is the potential difference across this resistor. Consequently all that is now required is high stability in the value of R_{fb}, a requirement which is easily satisfied. Basically sound design of this sort is typical of the author's electronic circuits, and I agree with him that there is no valid reason to suspect the overall reliability of the instruments with which they are associated.

From Fig. 10, it does seem that the author has been forced somewhat from his stated principles in so far as the overall behaviour of the circuit appears to depend directly upon the characteristics of the buffer and rectifier stages. It must be admitted, however, that it is not easy to see how this could have been avoided. The working conditions of the 85AI voltage reference tube also seem unusual. The burning current appears to be about 9mA, whereas the manufacturer's rated maximum current is 8mA, with a preferred value of 4·5mA. Is there perhaps some special reason for adopting such a large current?

Turning to more general matters, I feel particularly indebted to the author in at least two respects. First, he has drawn my attention to the way in which the instantaneous values of periodic quantities of supply or near-supply frequencies may be investigated by using a high-speed polarized relay. All too often one reaches for an oscillograph without pausing to think whether there are simpler and cheaper means available. Secondly, the author has introduced to me the Maxwell commutating bridge. This is so obviously a useful circuit that it is strange that it should seem to be so little known and used at the present time.

The author's simple and potentially reliable and accurate tachometer based on the Maxwell commutating bridge seems assured of wide application. It would also seem possible to work it from an electromagnetic pick-up device and avoid the necessity for a mechanical drive to an a.c. tacho-generator. If the pick-up generated one impulse per revolution, a maximum speed of no less than 12 000r.p.m. could be measured for a rate of relay operation of only 200 times per second.

In Section 3.2.2 it is not obvious why it is particularly convenient to make P the variable to effect range changing. It would seem equally convenient to change Q or S. Also, in Section 3.4, it is implied that the four ranges of the tachometer mentioned are varied by changing only P. May it not also be necessary to change C in order to ensure that it is always fully charged at each operation of the relay?

* ASTBURY, N. F.: "Industrial Magnetic Testing" (Institute of Physics, London, 1952).

BEFORE THE NORTH STAFFORDSHIRE SUB-CENTRE, AT STAFFORD, 3RD MAY, 1955

Mr. J. Wainwright: I should like to endorse a statement which the author made in his opening remarks on the need for the "engineering of measurements." Unless we have these periodic reminders of progress we are very apt to continue using instruments similar to those used in Clerk Maxwell's day. It is essential to "ruggedize" instruments, as the Americans put it, so that they are suitable for use in engineering shops. A good example of this is the electronic detector in place of a vibration galvanometer for a.c. bridges. This electronic device can be made with more than adequate sensitivity, and no difficult adjustments are needed to set up the instrument, whereas a vibration galvanometer can be very difficult in this respect.

The author states that electronics has been regarded with suspicion and this view still persists. A reading obtained from a pointer-type instrument will often be believed in preference to a reading from an electronic device. This raises the question of how the author applies checks to his instruments, and it would be interesting to know how often these checks are performed. The statement that the slip-meter "either reads correctly or not at all" is not entirely convincing.

I was rather surprised to see no references to the work of Koppelmann,* who, with his *Anglejoch* (whole-sheet type tester) and *Vektormesser*, has developed a mechanical-type rectifier with applications similar to those described in the paper.

The whole-sheet tester has been installed at nearly all the steelworks in this country, and although we do not wish to throw suspicion on the instrument, we have found no marked improvement in the quality of transformer sheet over the last 10 years.

I should now like to refer to dielectric measurements, and to ask what range and sensitivity of tan δ can be obtained on the author's instrument. Although it is agreed that bridge methods are not entirely satisfactory for rapid measurements in workshops, they are used to a considerable extent with great success. The Schering bridge, for example, introduced about 30 years ago, has remained a firm favourite, and will undoubtedly continue to do so. The biggest disadvantage is the need for a low-loss capacitor to withstand the full test-voltage, but since the author's device apparently needs frequent checking against such a standard, it is difficult to see what advantage there is, apart

* E.g. Instruments and Measurements Conference, Stockholm, 1949, p. 227.

from some time saving. If rapid measurements are required there is a great deal to be said for the recording Schering bridge as described by Geyger* and Keinath.†

In the paper it is inferred that Wagner earthing devices are needed with a.c. bridges, and although this may be true in certain cases in the laboratory, it is certainly not true in workshop testing. A further point on which I disagree with the author is in regard to the measurement of capacitance. It is stated that the measurement of capacitance is an unnecessary complication with the Schering bridge; in my opinion this measurement is not only useful but, in many cases, essential. For instance, during the impregnation of cellulosic materials with mineral oils, measurement of capacitance is the only way in which complete impregnation can be assured.

* GEYGER, W.: *Archiv für Elektrotechnik* 1937, 31, p. 115.
† KEINATH, G.: *Archiv für technisches Messen*, December, 1935, V.339.

Prof. F. A. Vick: I should like to ask the author whether he has tried using the Hall effect in germanium, or similar semi-conductors, as a wattmeter? While not so inherently stable as a dynamometer instrument, owing to temperature effects, etc., the semi-conductor instrument has the great advantage of a voltage output which is proportional to power input over a wide frequency range and which can be amplified in the normal way.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSIONS

Mr. D. Edmundson (*in reply*): I was particularly interested in Prof. Greig's opening remarks, for it was the knowledge of his pioneer work in the application of negative-feedback amplifiers to the solution of measurement problems that first stimulated my own interest, and that, I know, of many others, in this powerful weapon. Several other speakers referred to what might be called the philosophy of the use of electronics. Mr. Coombes has pointed out the rigorous conditions which must be satisfied, and others—Mr. Wainwright and Mr. Taylor—emphasized the importance of maintenance and standardization. The latter is as simple as it is important; owing to the negligible consumption of amplifier-operated instruments, they can be checked by the aid of standard impedances. Such arrangements should be invariably provided in permanent form.

Dr. Arnold refers to the use of amplifier-wattmeters for audio-frequency measurement, and while I agree with his remarks, I would emphasize that our instrument was developed for power-frequency work. The check at higher frequency is a useful criterion of phase-angle error. Although, as Mr. Ryder points out, the checking circuit refers only to zero and unity power-factor, in fact any combination can be used. With regard to the method of measurement at low power-factors, the mutual inductance used is—normally—a carefully designed toroid. For 50 c/s work, it is possible to apply correction for phase-angle errors, if these should be found on measurement. Its value is immaterial within fairly wide limits—in effect, those of satisfactory operation of the amplifier. As Mr. Ryder points out, care must be taken in ensuring that this is not overloaded. Although I have never used a Hall-effect wattmeter, I am in full agreement with Prof. Vick that this has great possibilities, in that it is our sole means of obtaining an electrical, instead of a mechanical, equivalent of power.

With Mr. Coombes's remarks about the versatility of Maxwell's commutator bridge I am in complete agreement. Not only can it be operated without loss of accuracy from electromagnetic or photo-electric pick-ups, but if the resulting frequency is too high for the relay, frequency-dividing circuits can be easily interposed. Its accuracy, however, is limited to about 0·1%: for the accuracy mentioned by Mr. Myers, counting methods are best; while it is not suited for the measurement of acceleration, as suggested by Prof. Greig. With regard to the details of the bridge circuit, queried by Mr. Coombes, the advantage in varying P lies in the fact that an equal deflection can be obtained on all ranges. C could alternatively be changed, but this (while theoretically better) is inconvenient. Mr. Wainwright queries the accuracy of the slip meter. This is particularly easy to check because, the coil being sensitive to the earth's field, a simple manual movement can be timed by a stop-watch.

Prof. Brailsford makes an interesting point with regard to the flux distribution in the whole-sheet tester. The slight dip in the curve is due to the impedance of the measuring apparatus connected in the central sections. The overlap between sheets does not affect the measurements, but is used because a gap occurring inside the coil would give rise to excessive primary current and would trip the protection. Mr. Taylor's remarks on single-sheet grading were also of great interest. He, with Mr. Langford, Mr. Martindale and others, refer to the limitations of my method in covering a variety of grades. I do not believe that single-sheet testing will ever be justified for low-silicon steel, so that in effect we are faced with variants of a single type of sheet. In contrast with the experience of Mr. Wainwright, we have found a substantial improvement in the quality of hot-rolled transformer sheet during the last ten years, by no means all of which can be accounted for, of course, by improvements in testing methods.

I am in general agreement with the remarks of Mr. Martindale and Mr. Rawlinson on the additional errors to be expected with high-permeability materials when making "a.c. permeability" measurements in a Lloyd–Fisher square, but these do not usually affect the argument in the one instance where routine permeability measurements are justifiable—low-silicon steel at high flux-densities. The Iliovici tester used was compared with N.P.L. and similar-type permeameters over the range of densities used, and agreed at least within $\pm 5\%$ in H. It only shows serious errors at very low or very high densities.

Mr. Edwards and Mr. Wainwright were critical of the method of measurement of the power factor of insulation samples. For the application for which it was designed—routine factory testing—it shows substantial advantages in time over the use of a Schering bridge, and provides a sufficient accuracy for a wide range of problems.

Finally, a number of speakers referred to the use of Carpenter's relay. This is not the place to provide detailed performance data, but it may be said that its outstanding merit is its complete freedom from contact bounce. It is to be hoped that before long this versatile device will take its place among accepted electrical measuring instruments.

621.395.625.3

Paper No. 1856 R
July 1955

A FLUX-SENSITIVE REPRODUCING HEAD FOR MAGNETIC RECORDING SYSTEMS

By E. D. DANIEL, M.A., Associate Member.

(*The paper was received* 20*th January*, 1955.)

SUMMARY

A description is given of a new type of flux-sensitive reproducing head for use on magnetic tape or drum recording systems. The head is compact and relatively simple in operation and differs from conventional reproducing heads in that its efficiency is fundamentally independent of signal frequency over the working range and is completely independent of the speed of the recording medium on reproduction. It can, in fact, be used to make measurements of recorded flux at points along the length of a stationary recorded medium. The head may be of value in certain specialized applications of magnetic recording technique, such as those dealing with very-low-frequency signals or the examination of transient phenomena.

(1) INTRODUCTION

A magnetically recorded signal may be regarded as a certain distribution of remanent magnetization along a suitable medium, such as a magnetically-coated plastic tape. The corresponding variation in flux along the length of the tape is in accordance with the time variation of the original signal, and in the ideal case the effective value of the recorded flux is independent of the frequency of the original signal. Such a recording is normally reproduced by moving the tape at constant speed past a reproducing head so designed that a proportion of the flux links with the coil of the head to generate an e.m.f. The e.m.f. so produced is proportional to the rate of change of flux along the tape and hence to the rate of change of the original signal with time. Fundamentally, therefore, the efficiency of a system using a conventional reproducing head is proportional to frequency.

In most audio-frequency applications of magnetic recording this feature of the reproducing head is not a serious disadvantage. Compensation for a 6 dB per octave rise in reproduced e.m.f. can usually be arranged in the reproducing amplifier down to the lowest frequency of interest without serious noise difficulties being encountered. When dealing with frequencies very much below the a.f. range, however, the inherent decrease in efficiency of the conventional reproducing head may impose a severe limitation on overall performance. To overcome this handicap, several magnetic recording systems designed for very-low-frequency work have employed amplitude or frequency modulation.[1] However, any form of modulation technique is extremely wasteful in tape since the tape speed necessary to accommodate the carrier and sidebands must be many times greater than that required to accommodate the highest frequency in the signal itself. In particular, such techniques would be difficult to apply in cases where the signal is not confined to very low frequencies, but contains frequencies over a large range extending down to a very low frequency. The tape speed required under these conditions might be prohibitively high.

A possible solution of some of these problems would be to redesign the reproducing head so that it responded to the instantaneous value of the recorded flux at the relevant point instead of to the rate of change of this flux with distance along the tape length. In this way the overall efficiency of a magnetic recording system would be made fundamentally independent of signal frequency. Moreover, by definition, the output obtained from such a head would be independent of the tape speed on reproduction, which might make it a useful tool for carrying out detailed examinations of recorded waveforms. If desired, point-to-point measurements could be made on a stationary tape.

So far, only one type of flux-sensitive head has been fully described.[2] In this, the tape flux is made to deflect an electron beam so that the voltage produced between two plates on which the beam impinges is a measure of the flux. The sensitivity obtained by this method is claimed to be high, but it suffers from the disadvantages of large physical size and, probably, of high cost compared with normal heads. Another method which has been investigated by the author and others[3] is based upon the principles of the second-harmonic type of magnetic modulator. Here, again, the head required tends to be rather large, since an extra magnetic circuit must be incorporated to balance the system and prevent the tape being erased by the exciting-signal flux. Also, with this principle, the whole of the main branches of the magnetic circuit must be taken to saturation before adequate sensitivity can be achieved. This, coupled with the fact that the excitation frequency must be in excess of the highest signal frequency, may make the power requirements of this type of head difficult to meet.

The flux-sensitive head to be described in the paper[4] uses the principle of a magnetometer described by Palmer.[5] It is similar, in some respects, to the second-harmonic modulator. The excitation frequency must be high, but only a small part of the core need be taken to saturation, and a transverse field effect is employed which obviates the need for a balanced magnetic circuit. The head may be made as small as any conventional reproducing head using similar core material.

The experimental investigation of the head has been carried out using magnetic recording tape. There is no reason, however, why the head could not be used on other forms of magnetic media, such as magnetically coated drums or discs.

(2) PRINCIPLE OF THE FLUX-SENSITIVE HEAD

As indicated in Fig. 1(*a*), the magnetic circuit of the flux-sensitive head is similar to that of a conventional reproducing head. During reproduction, flux from the tape is caused to flow round a suitably shaped core of low reluctance, rather than take

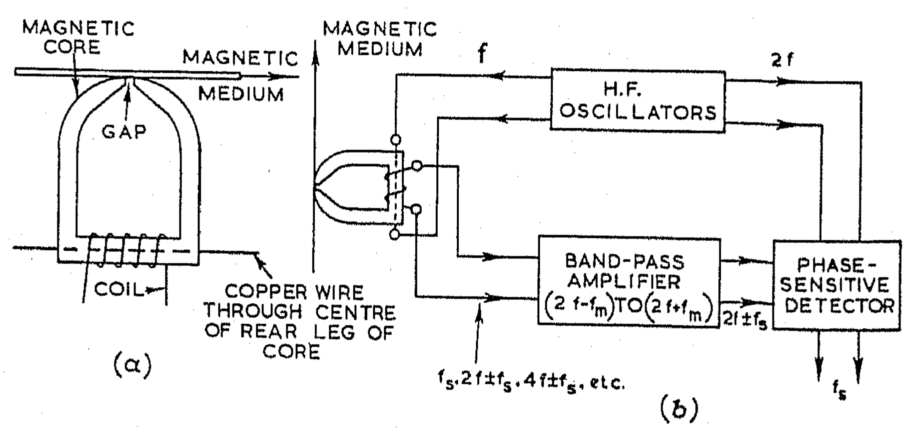


Fig. 1.—Basic construction of flux-sensitive head and associated electrical system.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Daniel is with the British Broadcasting Corporation.

its natural course, by the presence of the narrow gap in the magnetic circuit. This, as will be shown later, introduces a "gap loss" of exactly the same magnitude as would be present if the head were operated on conventional principles. Through the rear leg of the core a copper wire is passed, which must be insulated with thin sleeving if the core is of conducting material. The rear leg of the core also carries a coil which should, as nearly as possible, be symmetrically wound with respect to the wire.

In operation the wire is fed with an alternating current of frequency f in excess of the highest signal frequency to be considered, and of peak value sufficient to take a part of the core surrounding the wire close to magnetic saturation by virtue of the circular field created. If the coil and wire are symmetrically placed, there is no mutual inductance, and the passage of current through the wire should not produce a fundamental component of e.m.f. in the coil. Moreover, provided that the core is not magnetically polarized in any way, its magnetic characteristic should be symmetrical and no even-harmonic components of e.m.f. should be produced. However, if a magnetized tape is brought into contact with the front of the head, part of the recorded flux will flow round the core and the latter will become polarized. The magnetic characteristic of the core material will therefore no longer be symmetrical, and even-harmonic components of e.m.f. will be generated in the coil, of which the second, of frequency $2f$, is of major interest. If this component is selected and fed to a phase-sensitive detector, the detector output will be found to be closely proportional to the polarizing flux in the head core, and hence to the recorded flux from that part of the tape opposite the gap. When the tape is moved past the head, as in normal reproduction, the polarizing flux will vary and a modulated e.m.f. will be produced by the head. For example, if a sinusoidal signal of frequency f_s has been recorded on the tape, the reproduced e.m.f. will consist of sidebands of frequency $2f \pm f_s$, $4f \pm f_s$, etc., and with suitable filtering, the output of the phase-sensitive detector will be a copy of the original signal of frequency f_s. In practice, as indicated in Fig. 1(*b*), the detector is preceded by a band-pass filter accepting frequencies in the range $2f \pm f_m$, where f_m is the highest signal frequency under consideration. A signal amplifier follows the detector to bring the signal level up to the desired value.

The principle of the flux-sensitive head may be better understood from the following approximate analysis. For simplicity, it will be assumed that the core does not become completely saturated during operation, but that the flux density B in the core is related to the field strength H by an equation of the type

$$B = aH + bH^3 + cH^5 + \dots \qquad (1)$$

Let the excitation current in the wire be of the form

$$i = I \sin \omega t$$

where $\omega = 2\pi f$.

The circumferential field strength created by this current at a distance r from the axis of the wire can be written

$$h_y = H_y \sin \omega_1 t \qquad (2)$$

where $H_y = 2I/r$.

Let the longitudinal field strength corresponding to the polarizing flux from the tape be of the form

$$h_x = H_x \sin \omega_2 t \qquad (3)$$

where $\omega_2 = 2\pi f_s$.

If h_z is the resultant of h_x and h_y, the component of induction parallel to the axis of the coil is, from eqn. (1), given by

$$B_x = h_x/h_z(ah_z + bh_z^3 + ch_z^5 + \dots)$$
$$= h_x(a + bh_z^2 + ch_z^4 + \dots)$$

Now

$$h_x^2 + h_y^2 \simeq h_y^2$$

since the circumferential field strength is always very much greater than the longitudinal field strength created by the presence of the tape. To a very close approximation, therefore,

$$B_x = h_x(a + bh_y^2 + ch_y^4 + \dots) \qquad (4)$$

Substituting for h_y from eqn. (2), expanding, and considering only those terms involving the second harmonic of the excitation-current frequency

$$B_x = -h_x(bH_y^2/2 + cH_y^4/2 + \dots)\cos 2\omega t$$
$$= -h_x(2bI^2/r^2 + 8cI^4/r^4 + \dots)\cos 2\omega t \qquad (5)$$

The useful flux linking the coil is given by

$$\Phi = \int_{r_0}^{r_1} 2\pi r B_x dr$$

where r_0 is the radius of the wire and r_1 is the effective radius of cross-section of the rear leg of the head core. Substituting for B_x from eqn. (5),

$$\Phi = -2\pi h_x \int_{r_0}^{r_1} (2bI^2/r + 8cI^4/r^3 + \dots)\, dr \cos 2\omega t$$
$$= 2\pi h_x[2bI^2 \log (r_0/r_1) - 4cI^4/r_0^2 + \dots] \cos 2\omega t$$

or

$$\Phi = Ph_x \cos 2\omega_1 t \qquad (6)$$

where

$$P = b'I^2 + c'I^4 + \dots \qquad (7)$$

and the coefficients b', c', etc., depend upon the dimensions of the wire and core and the magnetic properties of the core material.

If the coil has N turns, the useful e.m.f. induced is given by

$$V = -N\frac{d\Phi}{dt}$$
$$= 2\omega_1 NPh_x \sin 2\omega_1 t$$

or writing $h_x = H_x \sin \omega_2 t$ and expanding

$$V = \omega NPH_x [\cos (2\omega_1 - \omega_2)t - \cos (2\omega_1 + \omega_2)t] \qquad (8)$$

i.e. the output consists of sidebands formed by modulation of the second harmonic of the exciting current by the signal, and its magnitude is proportional to the flux from the tape.

(3) DESCRIPTION OF EXPERIMENTAL SYSTEM

An experimental flux-sensitive reproducing system has been constructed, on the principles outlined in the preceding Section, which uses an excitation-current frequency of 50 kc/s and is capable of reproducing signal frequencies below about 10 kc/s. The head magnetic circuit is of ferrite material, which makes insulation of the wire through the rear leg unnecessary. The method of construction of the head is illustrated in Fig. 2. Two conventional ferrite reproducing-head cores were assembled in the normal way, and then fixed together with a suitable adhesive after grooves had been milled in the appropriate places on the rear legs of the cores. A copper wire of No. 36 s.w.g. was subsequently threaded through the channel formed by the mating of the two grooves, and a coil of approximately 50 turns was wound on the rear leg. The particular head used for making the tests to be described in the paper was clamped in an assembly designed for use with magnetic tape, but if desired, a head of exactly the same construction could easily be adapted to reproduce from a rotating magnetic drum or disc.

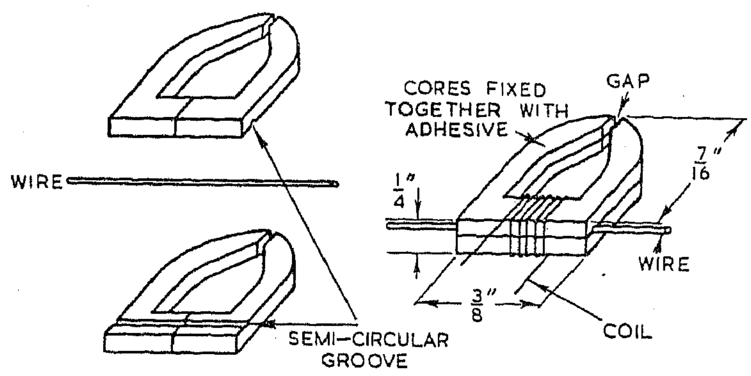


Fig. 2.—Construction of experimental head.

removed from the head, the resistance is adjusted until zero second harmonic of the excitation current is observed at the input of the demodulator. Normally the current required in the coil to achieve this condition is of the order of microamperes.

(4) PERFORMANCE OF THE EXPERIMENTAL SYSTEM

The sensitivity of the head is a function of the degree of non-linearity present in the magnetic characteristic of the material over the working range. Consequently, it is desirable that a high excitation current be used so that an appreciable part of the core is taken to saturation on the peaks of current. Fig. 4

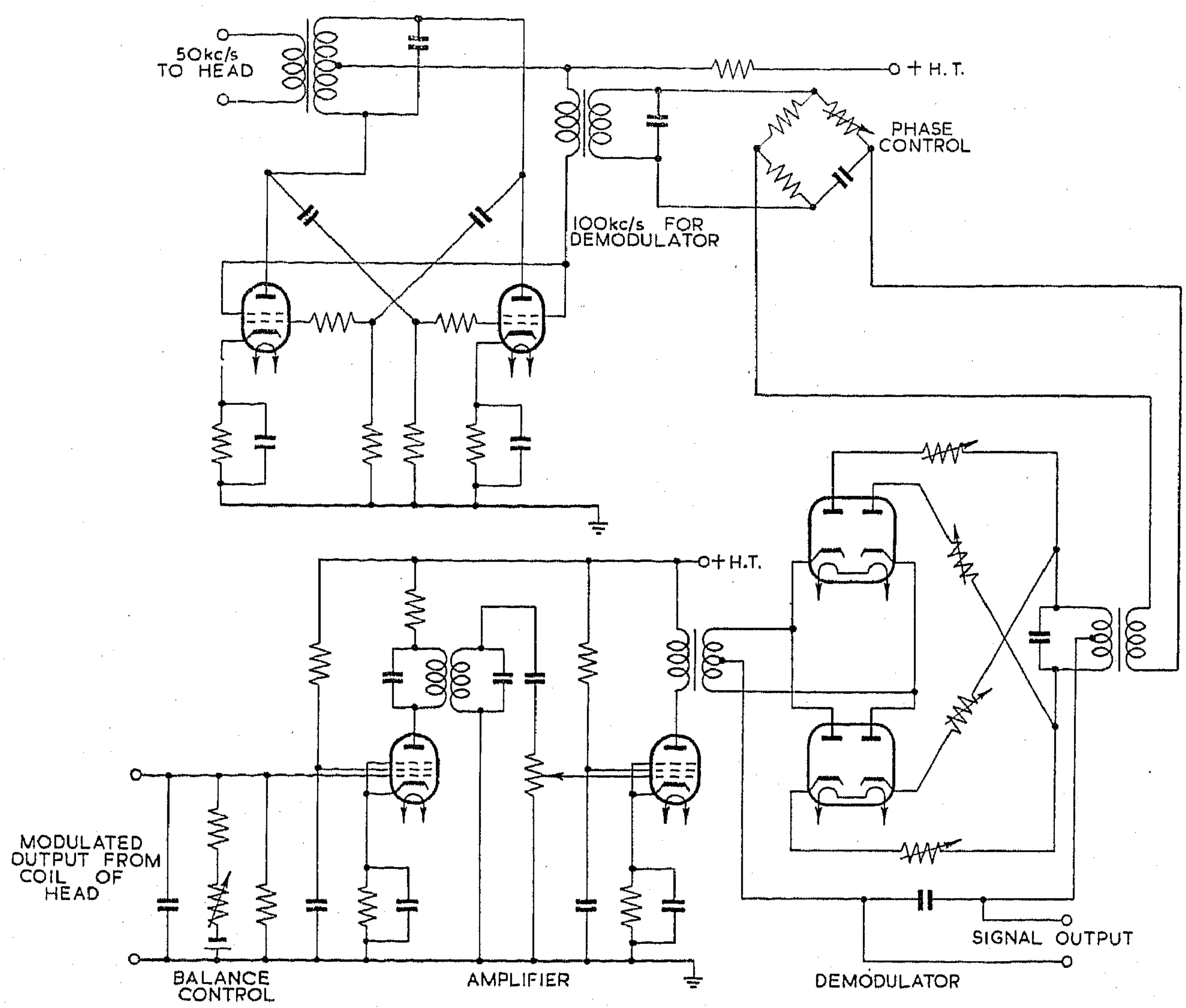


Fig. 3.—Electrical circuit of experimental reproducing system.

The electrical circuits associated with the head are shown in Fig. 3. The 50 kc/s excitation current is supplied from a multi-vibrator which also provides a 100 kc/s feed, through a phase-shifting network, to the balanced demodulator. The wire is tuned to 50 kc/s with a suitable condenser, and the coil is tuned to twice this frequency. The output from the coil is fed, through a stage passing frequencies in the range 90–110 kc/s to the demodulator, which is followed by a suitable signal-frequency amplifier. Measurements on a stationary tape require a d.c. signal amplifier unless, as is probable, only the magnitude of the flux is of interest. In that case measurements can be made on an oscilloscope connected before the demodulator.

In practice, the head is found to be extremely sensitive to the presence of external polarizing magnetic fields, even though it is screened by a double-walled Mumetal box. Also, since the core material has a comparatively high retentivity, trouble may be experienced owing to remanent magnetization in the core. To overcome these difficulties a balancing circut has been introduced. As shown in the Figure, this consists of a small battery feeding the head coil through a large variable resistor. With the tape shows the output obtained at various excitation currents, using a signal frequency of 100 c/s and a tape speed of 15 in/sec. The phase of the head output changes with excitation current, but a corresponding phase adjustment can be made in the demodulator as each new value of current is selected. It is clear that, at the

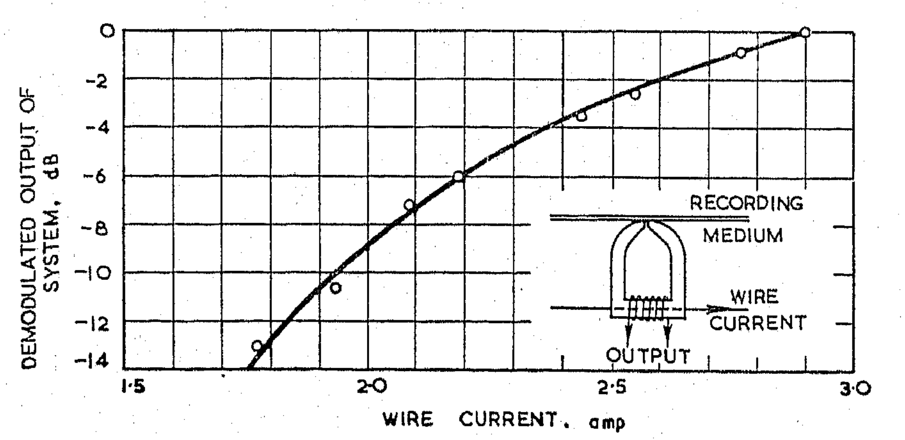


Fig. 4.—Variation of sensitivity with the excitation current in the wire.

maximum value of 2·9 amp available, the sensitivity has not yet reached its maximum, although the curve has begun to rise less steeply. The remainder of the experiments to be described were carried out using the maximum excitation current available.

Frequency responses of the head were measured at a variety of tape speeds from $7\frac{1}{2}$ to 100 in/sec, and from the results obtained (denoted by crosses) the composite curve of Fig. 5 was plotted. In this Figure, the output of the head, corrected for electrical losses in the recording and reproducing chains, is plotted against the reciprocal of the recorded wavelength. The results are therefore in the form of a universal curve, from which the response/frequency characteristic at any particular tape speed can easily be deduced.

The fall in the response of the head at short wavelengths is attributable to the various losses that occur during the recording process, and to the so-called "gap loss" in the head itself.[6] This was demonstrated by using the head as a conventional reproducing head. The 50 kc/s excitation was disconnected, and the coil output was fed into an integrating reproducing amplifier to correct for the fundamental 6 dB/octave rise in response. The relative output obtained under these conditions is denoted by the circled points in Fig. 5. The slope of the curve in the medium-

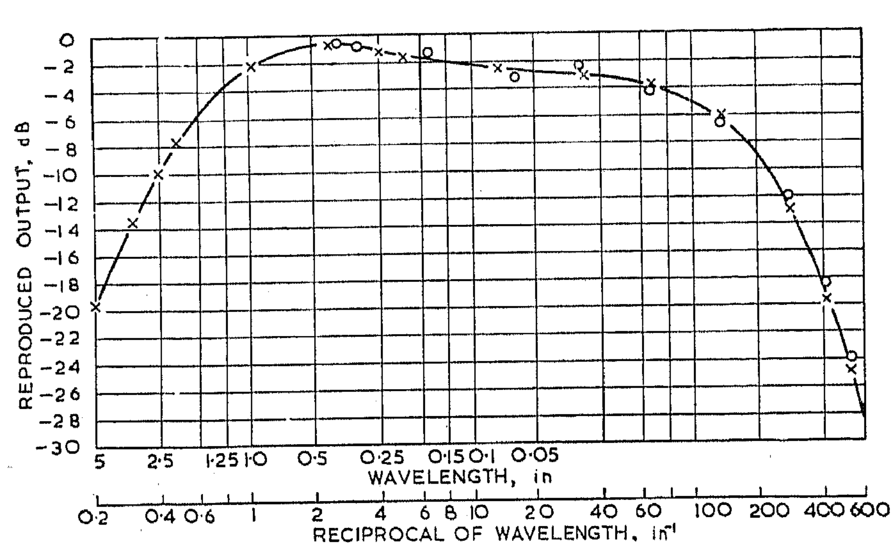


Fig. 5.—Response of the experimental head as a function of the reciprocal of the recorded wavelength.

× Used as a flux-sensitive head.
○ Used as a conventional reproducing head and corrected by 6 dB/octave.

wavelength range (theoretically the slope should be zero) is also a phenomenon found in conventional reproducing heads, and this is again demonstrated by the circled points. In fact, the only difference in the response of the head under the two conditions of operation is the expected one of 6 dB/octave between the two frequency characteristics.

The decrease in the response of the flux-sensitive head at very long wavelengths is a function of the overall dimensions of the head. For the response to remain flat the pole-pieces of the head must be capable of collecting all the flux available from at least half a wavelength of the tape. Obviously, as the recorded wavelength becomes larger, the tape flux spreads further out from the tape surface so that this condition cannot be achieved. Very approximately, it might be expected that the response should fall by 6 dB when the overall dimensions of the head are equal to a quarter-wavelength. In the present instance, these dimensions are of the order of $\frac{3}{8}$ in (approximately one-third of those of a conventional iron-cored head) so that a 6 dB loss should occur at a wavelength of about 1·5 in. An examination of the curve of Fig. 5 shows this to be approximately correct. Obtaining an improved long-wavelength response does not necessarily entail enlarging the whole head. It is sufficient to lengthen the front surface of the core by, for instance, mounting high-permeability metal flanges which extend outwards each side in proximity to the tape.

The signal/noise ratio obtainable from the system in its present form is about 40 dB relative to the maximum signal output for 3% total harmonic distortion. This figure was actually measured at a tape speed of 15 in/sec with the long-wavelength response left as indicated in Fig. 5, but with the high-frequency response equalized to be flat up to 10 kc/s. A signal/noise ratio of 40 dB is less than would be obtained from a conventional a.f. recording machine at the same tape speed. This is partly to be expected, in that the flux-sensitive system has a considerably better response at long wavelengths, and in this region, slow variations in the magnetic properties and dimensions of the coating along the length of the tape may give rise to appreciable noise components. A large part of the noise from the experimental system is not, however, fundamental and could probably be considerably reduced in equipment designed for a specific application.

The upper limit to the signal frequency which could be accommodated by a head of this type depends primarily upon the highest excitation frequency that can be used. The impedance characteristics of the excitation circuit of the experimental head, over the frequency range 50 kc/s–1 Mc/s, are shown in Fig. 6.

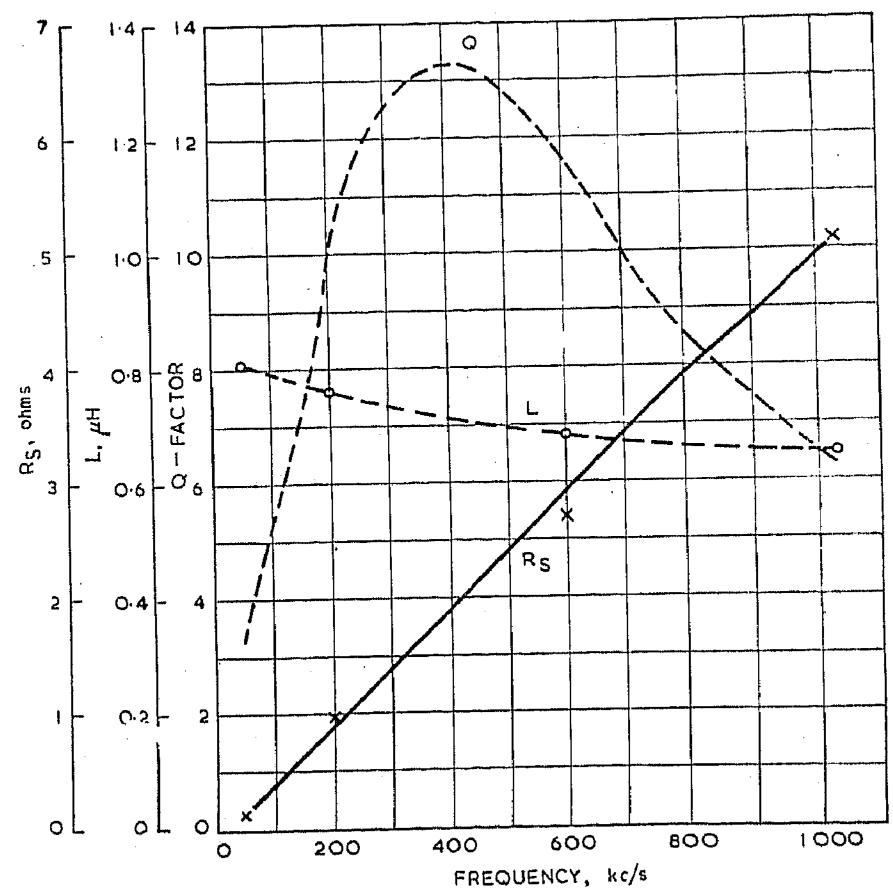


Fig. 6.—Impedance characteristics of the head at various excitation-current frequencies.

These results were obtained at low currents, but they give an idea of the way in which the impedance of the head increases with frequency. They indicate that the wire has an inductance of approximately 7 μH and an effective series resistance which is proportional to frequency. At 50 kc/s the resistance is approximately 0·1 ohm, so that the power required to provide the necessary excitation at this frequency is about 1 watt. In practice, hysteresis losses at the high flux densities involved make the power requirements considerably higher. At excitation frequencies much above 200 kc/s it might be difficult to provide sufficient excitation without undue heating of the core. This limit could, however, be raised by using a different grade of core material and by feeding the wire with a non-sinusoidal current of high peak value but low r.m.s. value.[5]

(5) POSSIBLE APPLICATIONS OF FLUX-SENSITIVE HEADS

The most obvious application of the flux-sensitive method of reproduction is where the signals are confined to very low frequencies and a long playing time is required. Conditions of this kind often occur in physiology and medicine when heart-rate or brain-wave phenomena are to be examined. The signals involved contain components in the range 1–100 c/s, and it is usually necessary to record for very long periods in order to observe a possible abnormality. The flux-sensitive head described in the previous Sections could meet such requirements quite easily. For instance, a substantially flat response from 1 to 100 c/s could be obtained without any equalization by running the tape at 1 in/sec. This is clear from Fig. 5, in which, at a tape speed of 1 in/sec, the abscissae represent frequency in cycles per second. Under such conditions there would, of course, be no need to use such a high excitation frequency. Also, it would be an advantage to use a much lengthened head core. In this way the extreme low-frequency response would be considerably improved, and the tape speed could be reduced to provide an even better playing time per reel.

The reproduction of signals such as pulses, which contain components over a large frequency range down to zero frequency, presents a number of problems. In the first place, a flux-sensitive head, in common with conventional reproducing heads, relies primarily for its operation upon intercepting as much as possible of the external tape flux. On these grounds, therefore, the reproduction of zero-frequency, or infinite-wavelength, signals is impossible, and most pulse-reproducing systems will have to incorporate some form of d.c. restoration. Nevertheless, the use of a flux-sensitive head may improve the situation at the very-low-frequency end in two ways—by increasing the response, and by providing a modulated output from the head and thereby easing some of the problems of transformer and amplifier design. Whether a flux-sensitive head of the type described can be used will, however, depend upon the maximum signal frequency to be reproduced, since, if this is very high, the excitation frequency required may exceed the practical limit. In some cases, the difficulties at high frequencies might be overcome by winding balanced coils on the two side limbs of a flux-sensitive head core and covering the high-frequency end of the range by conventional reproducing technique.

A further application of the flux-sensitive head may arise in cases where a detailed examination of various recorded information, such as transient phenomena, is required. The recorded information could be replayed at very slow speed, or, if necessary, discrete measurements of the information could be made at various points along the tape. The method would have advantages over the normal pen-recorder technique in that very rapidly varying transients could be accommodated by adjusting the tape speed on recording, and long-period analyses could be made without wastage of recording material.

(6) ACKNOWLEDGMENTS

The author wishes to acknowledge the valuable part played by Mr. W. T. Frost in carrying out the experimental work. The paper is published by kind permission of the Chief Engineer of the British Broadcasting Corporation.

(7) REFERENCES

(1) MOLYNEUX, L.: "Recording Low Frequency Phenomena on Magnetic Tape," *Electronic Engineering*, 1952, **24**, p. 130.

(2) SKELLET, A. M., LEVERIDGE, L. E., and GRATIAN, J. W.: "Electron-Beam Head for Magnetic Tape Playback," *Electronics*, October, 1953, p. 168.

(3) United States Patent No. 33238, 1949.

(4) British Patent Specification No. 10565, April, 1954.

(5) PALMER, T. M.: "A Small Sensitive Magnetometer," *Proceedings I.E.E.*, Paper No. 1551 M, October, 1953 (**100**, Part II, p. 545).

(6) DANIEL, E. D., and AXON, P. E.: "The Reproduction of Signals recorded on Magnetic Tape," *ibid.*, Paper No. 1499 R, May, 1953 (**100**, Part III, p. 157).

DISCUSSION ON "SOME APPLICATIONS OF THE ELECTROLYTIC TANK TO ENGINEERING DESIGN PROBLEMS"*

Before the NORTH-EASTERN RADIO AND MEASUREMENTS GROUP *at* NEWCASTLE UPON TYNE, 1*st March, the* NORTH-WESTERN CENTRE *at* MANCHESTER, 4*th May, and the* SOUTH-WEST SCOTLAND SUB-CENTRE, *at* GLASGOW, 1*st December*, 1954.

Mr. J. Wainwright (*communicated*): The paper mentions only solutions which have been dealt with adequately in the literature, and there is no reference to three-dimensional fields having no planes of symmetry. Although the typical examples given in the paper probably cover the majority of engineering design problems, the need occasionally arises for investigations into the more complex fields. Such investigations can be carried out only in a deep tank containing a three-dimensional model and using a probe which can be adjusted in all directions. Published information on this type of measurement appears to be very scarce. Peplow† has given some very useful practical information on the capacitive type fields which are found in high-voltage switchgear installations. Schmidl,† in addition to general information on the use of tanks, has given some details on "mixed" fields of the capacitive-resistive type. Have the authors any information or experience concerning this type of measurement?

Mr. C. G. Giles (*at Newcastle upon Tyne*): Although the electrolytic tank as an analogue method of solving field problems has been known since about 1875, it really became a practical tool in the mid 1920's with the development of electronic oscillators, amplifiers and detectors. The field of application to the tank was greatly increased in the 1940's by the methods suggested for representing dielectrics of different permittivity by stepping the tank bottom and the use of conducting pins.

I have used the technique since 1928 for the solution of a wide

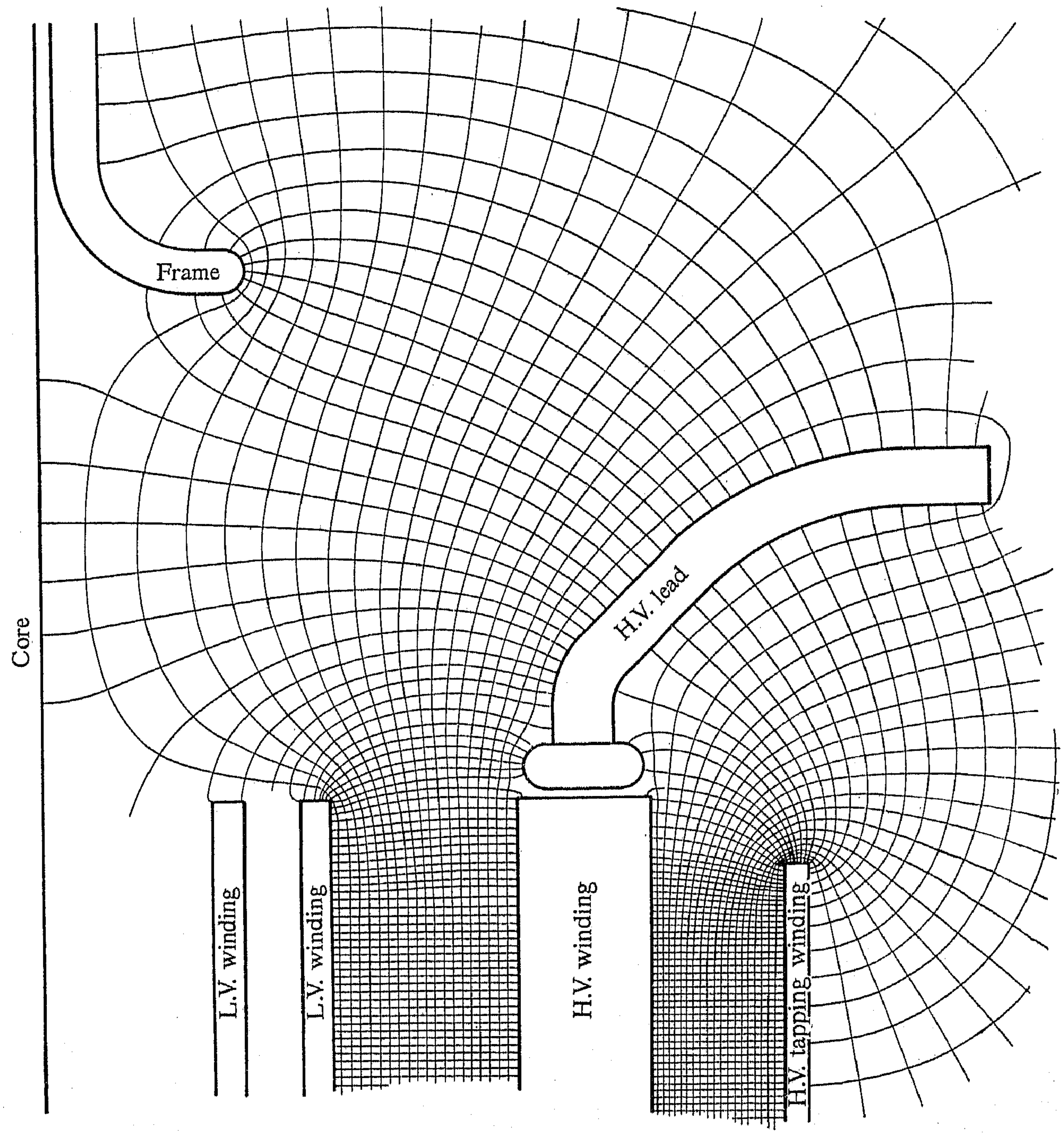


Fig. D.—Electrolytic-tank plot of equipotential lines in a section of a transformer.

* DIGGLE, H., and HARTILL, E. R.: Paper No. 1267 M, February, 1954 (see 101, Part II, p. 349).

† PEPLOW, M. E.: *Electrical Times*, 15th February, 1951, p. 256.

† SCHMIDL, H.: *Elektrotechnik und Maschinenbau*, 1953, 14, p. 309.

range of problems in transformer and alternator design relating to the distribution of electric stress in the vicinity of windings and bushings, magnetic-leakage field plots and thermal gradients in stator- and rotor-winding assemblies. In addition, investigations were made on turbine-blade assemblies to determine the effect of alteration in blade contours and different inlet and outlet angles.

Fig. D shows the measured equipotential lines and freehand plot of lines of equal stress to conform with an orthogonal configuration of a longitudinal section of the top part of the windings of a transformer, including the main lead from the h.t. winding, as obtained in an electrolytic tank.

In order to permit a straightforward experiment to be carried out it was assumed

(i) That the radii of the core and coils were sufficiently large in relation to the section of the transformer under tests, so that a 2-dimensional plot could be made on a longitudinal plane through the axis of the coils.

(ii) That the insulation between windings and core was homogeneous.

(iii) That the transformer-tank walls were sufficiently far away from the region under test to have negligible effect.

Fig. E shows an application of the electrolytic tank to mechanical engineering. This is a representation of the streamline flow around one of a cascade of turbine blades under certain specified conditions.

If a conducting body is immersed in a fluid between two charged plates, the equipotential lines in the fluid surrounding the immersed body are exactly similar to the lines of flow passing around the body, or through which the body is moving with a velocity parallel to the charged plates. Under these conditions, there can be no non-parallel component produced as the result of simple flow in a perfect fluid, and in order that an effect analogous to the "lift" of a blade may be obtained, it is necessary to superimpose upon the perfect fluid streamlines a circulation about the body. The effect of circulation is given by the application of potential to the blade itself—in this example, 45% of the upper blade potential.

At the left- and right-hand sides of the diagram the equally-spaced equipotential lines represent equal quantities of fluid moving in parallel lines and with the same velocity in the direction of the blade. As the blade is approached, the lines divide and pass around the blade, more passing around the upper than the lower surface. Since the gradient above the blade is steeper and more fluid has to take this path, the velocity of the fluid over the upper face will be greater than that over the lower. From Bernoulli's theorem, the total energy of a stream of fluid is constant when no work is done. Since the fluid on the upper surface of the blade increases in velocity its pressure must decrease, whilst conversely, the pressure of the fluid passing below the blade must increase. There is thus a component of pressure on the lower surface of the blade. It is this unbalanced pressure component which gives the blade its "lift."

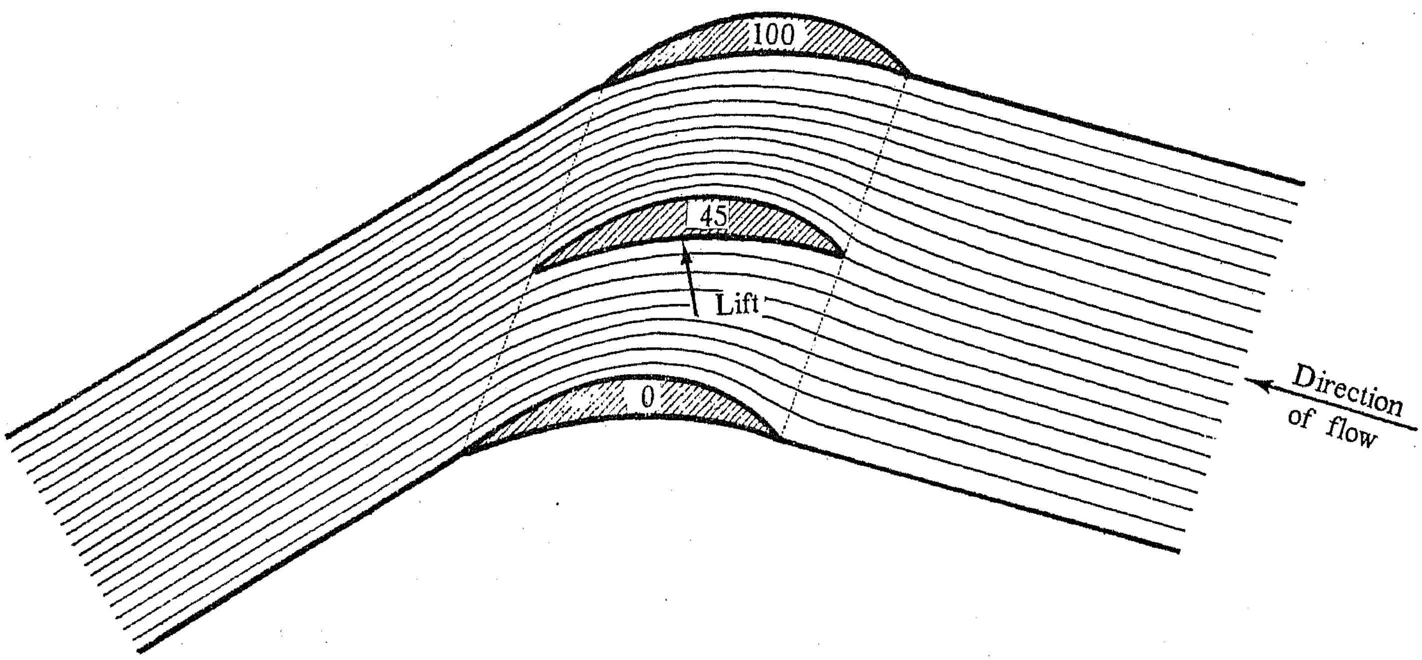


Fig. E.—Electrolytic-tank plot of blade streamlines.

Mr. R. B. Burtt (*at Newcastle upon Tyne*): There is no mention in the paper of the electrode materials that were used in the tank, nor of the type of probe used. Do the authors feel that the hard water of the north-east would be as suitable for field mapping as that supplied by the Manchester Corporation?

It is interesting to note that the authors use an approximation in deciding the width of the slot in Fig. 5 that represents the porcelain shed. Have they thought of using some sort of Carter's gap coefficient in deciding the depth of the slot, in order to avoid the trouble of pinning the boundaries? Have the authors had any experience in the use of solid semi-conductors for simulating materials of high permittivity?

Electrolytic-tank models are often used for the determination of the potential grading of the elements in large circuit-breakers. The model is usually a 3-dimensional one and of a relatively small scale. Approximations are used to a large extent, in that resistors represent the lumped capacitances of bushings, and conductors are thickened where necessary to allow for adjacent porcelain masses. Results obtained with this type of model have agreed very well with full-scale test values. I would be interested to know the authors' experiences of this kind of tank work.

Mr. J. M. Elliott (*at Newcastle upon Tyne*): The authors' method of recording the equipotentials is especially good. It is advantageous to be able to see the form that the equipotential pattern is taking whilst carrying out the actual plot, and I am sure that the cross-wire indication directly above the probe will enable the time spent in actual experiment to be reduced. Whereas all the other systems do not necessarily entail inaccuracies, by virtue of their recording methods, that used by the authors is compact and must therefore be a useful addition to any research or design department. The method of calibrating the tank using concentric electrodes is ideal, since the limitations normally imposed by the tank boundaries are entirely eliminated; these effects were well indicated in a tank I constructed some years ago. It was decided to check a known field, and as is usual when checking a device which relies on graphical plotting,

one containing straight lines and circles was considered desirable. The field surrounding two parallel conductors was therefore chosen. This consists of circles surrounding the conductors with their centres on the projected line joining the centres of the conductor, but compressed towards the space outside the conductors, and one straight line midway between the two conductors and at right-angles to the centre line. On plotting the field from a model in the tank, this was found to be the case, and true circles were plotted immediately around the conductor while the mid-potential line was straight. Towards the outside of the plot, however, the field was compressed inwards, thus behaving as though there were other conductors parallel to the first pair but outside the tank, and this reflected image is the effect of the tank boundaries. The effect was apparent on all four walls, but was more noticeable at the sides, owing to the dimensions of the tank. Thus as the tank boundaries take up a potential and influence the field, the models have to be suitably arranged to take this into account, as mentioned by the authors.

Some interesting plots were also taken, showing the electrostatic fields existing in electronic valves, although they would not be directly applicable when the valves were conducting. One particular plot depicted the field inside a variable-mu valve, and may be of use to lecturers in electronics. It showed how the valve is made to cut off gradually, by virtue of the variable grid spacing affecting the local intensity of the field. The model used was that of a valve having a spiral grid, the mesh being fine at one end and opening out to a coarser mesh at the other.

Mr. A. L. Shaw (*at Newcastle upon Tyne*): In the Introduction, there is a reference to "the well-known technique of curvilinear squares." My experience is that very few engineers engaged in the solution of general electrical engineering problems (as opposed to those of specialized design) are, in fact, conversant with this method of field mapping. Knowledge of the method is not enough; as stated by Attwood* "facility in field mapping is acquired only by hours of persistent endeavour expended in the solution of field problems." This was also made clear by Moore,† who has given a graphic description of the method. I mention this because the non-specialist engineer with a problem of field plotting to solve often wastes time in looking up details of the method only to find that, by the time he has acquired the necessary practice, it would have been quicker and cheaper to obtain the use of an electrolytic tank.

The authors have not mentioned one application of the electrolytic tank to which I think it is particularly suited, namely the design of liquid rheostats, particularly water rheostats, for the artificial loading of large water, steam or gas turbo-generator sets.

The electrolytic tank as described by the authors is essentially a 2-dimensional field plotter. An improvement could be made by using a conducting gel as the electrolyte and by colouring this gel with a thermometric dye having a suitable number of irreversible colour changes over a range of temperatures from, say, 60 to 200°F. By passing a relatively high current through a tank of this nature for a brief period it should be possible to produce a self-plotted field in three dimensions simultaneously, based on the resistive losses between the various equipotential lines. The gel electrolyte could then be sliced into suitable sections, and the results would be comparable to those obtainable by the photo-elastic method of mechanical stress analysis. This technique would, of course, depend on the availability of a suitable gel and also of a thermometric dye giving a number of colour changes at relatively low temperatures. Typical thermometric pigments available commercially at present can give four colours in succession within the range 100–700°C, which is inconveniently high. However, manufacturers of such pigments or dyestuffs might be able to develop a suitable product if the proposed application were brought to their notice.

In Section 3.3.1 mention is made of the study of heat flow through the walls of buildings, and it would be interesting to have an illustration of such a problem and its solution. The method of reproducing the temperature gradient in the air adjacent to flat-wall surfaces would be of particular interest. A common problem in the heating of buildings, which would lend itself very well to solution by the electrolytic tank, is the distribution of floor-surface temperatures for various installed depths and lateral spacings of embedded floor-heating elements. Such elements commonly take the form either of hot-water pipes or electric resistance cables, the latter being an increasingly popular method of heating large buildings. With rooms having relatively high heat losses (and therefore high heat loading per square foot of floor), it becomes important to know just how deeply to embed the heating cable in order to avoid uncomfortably high surface temperature, and at the same time to avoid undue increase in floor-screed thickness. It is also useful to be able to estimate the temperature of the wall and floor adjacent to the skirting boards, since this is, in some cases, a convenient location for the supply wiring.

Dr. R. Feinberg (*at Manchester*): The designer of new equipment has fundamentally the choice of three different methods of approach to enable him to find the required design data. First, he may use the method of extrapolation from already existing similar designs—in other words, he may make an intelligent guess; secondly, he can establish the data by mathematical treatment of his problem; and thirdly, he can find the data by measurement on suitable model designs. The authors have shown a number of practical engineering problems where the third method is the appropriate line of approach in order to achieve a satisfactory result.

With regard to the example given in Figs. 2, 3 and 4, where the dielectric components in the electric field are air and porcelain, the respective permittivities are 1 and 6 to 8, and the higher permittivity of the porcelain is simulated by the authors in their electrolytic tank by an increased depth of the electrolyte. How far is it accurate to apply this method of simulation, and is there not a substantial factor of distortion introduced in the simulation of the electric-field conditions?

Would the authors express an opinion on the use of the low-conductivity paper method, instead of the electrolytic tank, for the mapping of electric fields?

Mr. W. P. Baker (*at Manchester*): The electrolytic tank has been subjected to some criticism over the past year or two on the score of accuracy, and it is worth commenting on the probable accuracy with which some of the boundary conditions may be specified. The Table in Section 3.3.2 of the paper shows how the permittivity changes with change of direction of stress, and Fig. F shows how the permittivity along the laminae varies with temperature for the same board. More important than the different values of permittivity along and through the laminae is the steepness of the permittivity/temperature characteristic between 80°C and 100°C.

Clearly a change in the operating temperature of a few degrees Celsius in this range will require a new solution. Although this type of board contains more resin than would be expected in a bushing, it might well be used as a tap-changer panel and be subjected to stresses along the laminae. If a temperature gradient is possible in the board, the permittivity will change with position and also with stress, and although from an engineering-design point of view the differences have their effect in the right way in relieving points of high stress, the example does illustrate that the accuracy of the tank is better than some of

* Attwood, S. S.: "Electric and Magnetic Fields" (Wiley and Sons, 1941).
† Moore, A. D.: "Fundamentals of Electrical Design" (McGraw-Hill, 1927).

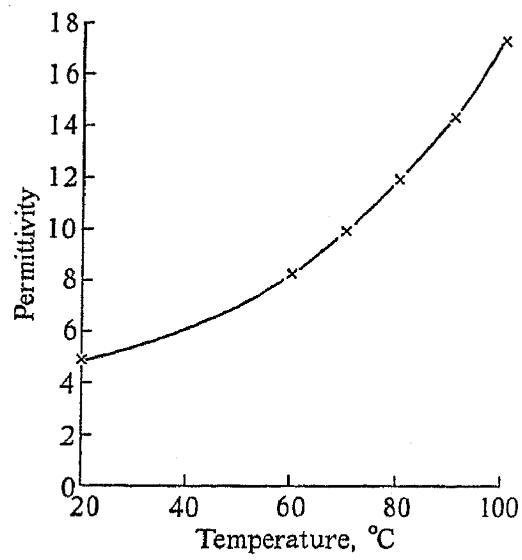


Fig. F.—Variation of the permittivity of s.r.v.p. board (stressed along the laminae) with temperature.

the information with which to start a problem. A few examples of magnetic materials might equally well be invoked to support the argument. On the few occasions that high accuracy is desirable and possible, a great saving in time may be effected by obtaining a good approximation by means of the tank and then relaxing the solution to the desired degree of accuracy.

Mr. V. H. Attree (*at Manchester*): I would like to describe a new type of probe assembly and detector unit which simplifies the actual plotting procedure. On the earlier tanks it was usual to employ a pantograph system which followed the path of the probe electrode so that the potential plot was obtained on a sheet of paper at the side of the tank. More recently the authors have used a horizontal sheet of Perspex located vertically above the surface of the electrolyte and having a lead-screw drive to the probe electrode. The plot is made on tracing paper placed over the Perspex sheet, and the exact position of the probe is indicated from below by illuminated cross-wires, thus avoiding the lost motion inherent in the pantograph system. Balance was determined aurally by means of telephones.

Prof. Bradshaw* has suggested that the plotting process could be speeded up if the probe were suspended by a permanent-magnet system having one set of magnets below the Perspex and the other set above. The probe could then be moved in any direction by moving the upper magnet assembly. This arrangement dispenses with the lead screws and gives complete freedom of movement to the probe. Having simplified the probe system, the next step is to eliminate the telephones. This is done by providing a visual indication of null balance directly on the light which illuminates the cross-wires. The "bright up" indication is obtained from a small electronic unit containing three valves. This unit amplifies and rectifies the unbalance signal and uses the resulting direct current to control an oscillator which operates the lamp. With an exact null indication the lamp is at full brightness, whilst with an unbalance of 1 mV (r.m.s.) the lamp is just extinguished. This sensitivity is adequate for all practical purposes, since it implies a discrimination in potential of 0·1% with an excitation of only 1 volt across the tank.

The arrangement works quite successfully and has been used for a number of potential plots. However, it has one disadvantage over telephones in that it is necessary to adjust the gain control occasionally in order to suit the potential gradient on the contour being followed. If this is not done, too high a sensitivity makes it difficult to locate the null point, whereas too low a sensitivity gives an uncertain null indication. The difficulty does not arise with telephones owing to the logarithmic nature of the amplitude response of the ear. It is hoped to eliminate the need for gain control by a new electronic detector unit having a logarithmic response.

The magnetic probe suspension associated with a really simple visual indicator appears to be much more attractive than a fully automatic plotter, which is complex and expensive.

Mr. J. D. Hodgetts (*at Manchester*): Fig. 15(*b*) is particularly helpful in the understanding of the particular feature of the coreless reactor, namely the straight-line voltage/current characteristic.

As is well known, any current-limiting reactor, particularly when it is used in a system with a high over-current fault rating, must have a well-defined reactance at maximum fault current. To this end a straight-line voltage/current characteristic is undoubtedly desirable, and Fig. 15(*b*) illustrates how this is achieved. The effectiveness of the cylindrical shield in confining the flux to a non-magnetic path (the requirement for constant reactance) is readily evident, and although in practice the steel tank is in close proximity to the outside of the shield, negligible flux passes into it. The ratios of the governing dimensions of the model approximate very closely to those used in modern practice, and the plot is therefore typical of most reactors at present in service or manufacture.

A reactor of similar dimensions to those of the model was short-circuit tested. The peak asymmetric short-circuit current was approximately 50 times the normal full-load r.m.s. value. Twenty-two tests were made and oscillographic records were taken. The reactance under asymmetric short-circuit conditions varied about 5% above and below the mean value, and the mean value showed no reduction from that at normal full-load current.

The hyperbolic base appears to be fundamental for a coil whose field path is entirely non-magnetic, and the authors' method of representation of the shield in the model seems to be in conformity with the electromagnetic and physical principles governing the relationship of the shield to the coil in practice. The slight errors which may occur owing to the longer current paths through the well of the hyperbola and the introduction of the

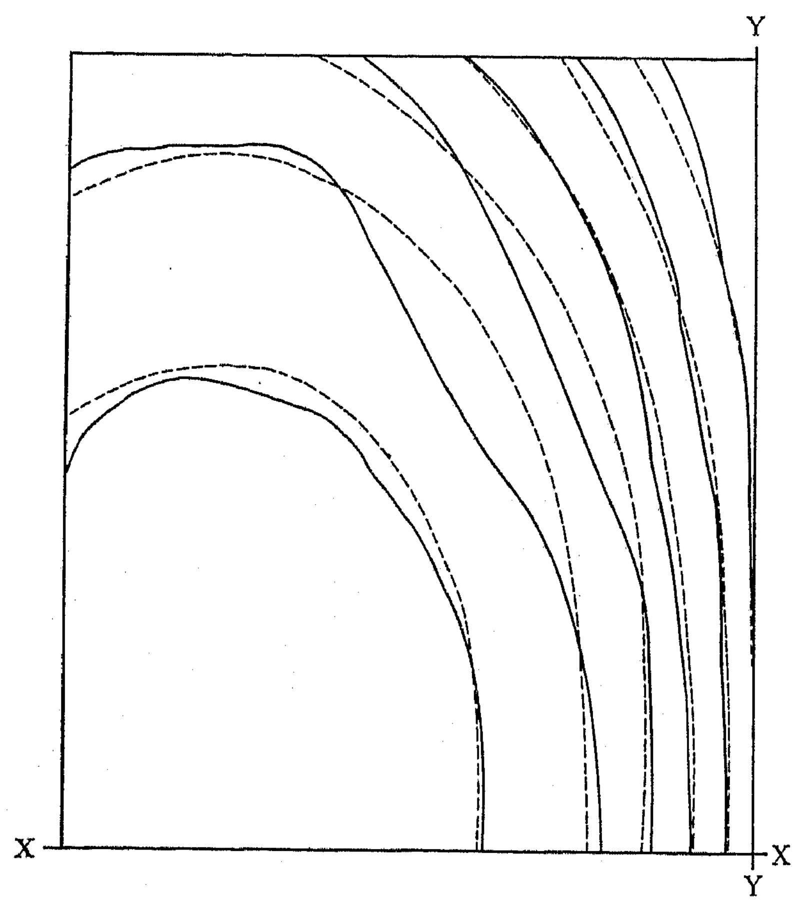


Fig. G

—— Electrolytic bath.
---- Measured field strength.
Y–Y Inside surface of coil.
X–X Axial centre line through coil.

* Bradshaw, E.: *Proceedings I.E.E.*, 1954, **101**, Part II, p. 364.

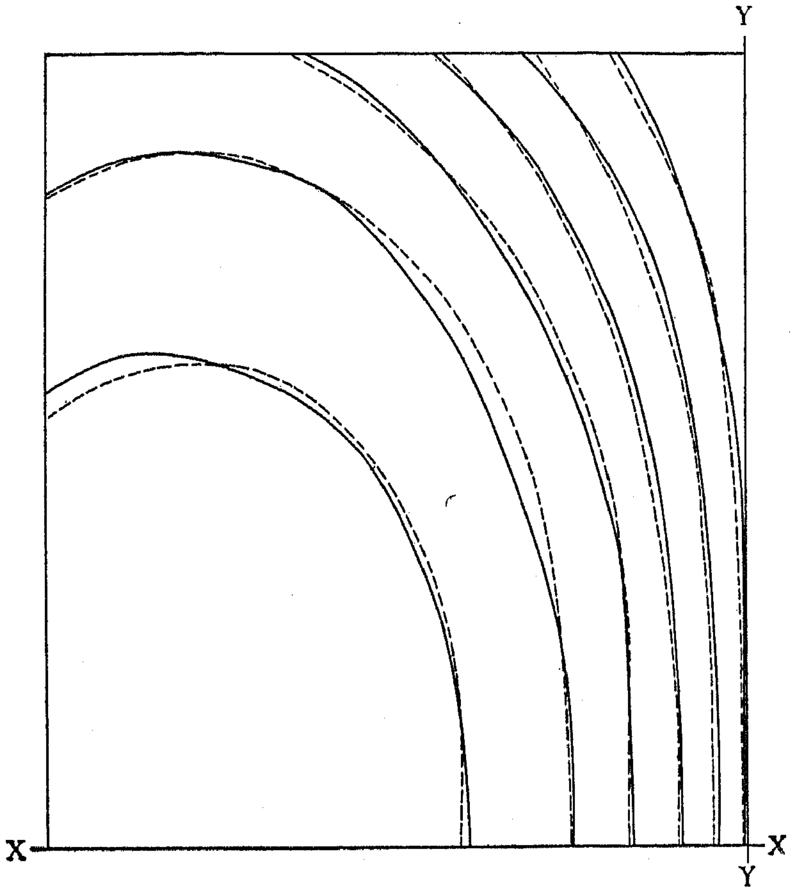


Fig. H

—— Calculated field strength.
---- Measured field strength.
Y–Y Inside surface of coil.
X–X Axial centre line through coil.

brass strip, are small and of little significance in the determination of flux density within the coil itself.

Equal resistances have been fitted to the model as shown in Fig. 16. Even so, it would be difficult to determine by calculation the actual current flowing through the individual probes representing the turns of the coil, although there may be some mutual adjustment of the length of the current paths and current densities in the electrolyte. The establishment of equal currents in the probes is certainly necessary for the accuracy of the field plot.

The comparison with Fig. 17 is interesting and indicates a method of determining local and r.m.s. values for the ultimate purposes of economic design.

Reference is made to the comparison of the field configuration given by the electrolytic bath and by means of a small search coil. Fig. G shows this comparison, and Fig. H shows the comparison between the field configuration given by the search-coil method and as given by calculation.

Mr. E. G. Wright (*at Manchester*): The electrolytic tank enables solutions of electric-field problems to be simply and quickly obtained, and it may be assumed that the technique is now well established. However, the information is not efficiently used if one is satisfied, as the authors appear to be, with a mere inspection of the field plots. Efforts should now be directed to determining the stresses which cause corona or breakdown in various dielectrics, in order that the information from the plots can be fully and effectively used.

Despite the complicated nature of the breakdown mechanism, it is already possible in some cases to predict breakdown voltages from field plots, and as more information becomes available, the electrolytic tank will become an indispensable tool for all designers of high-voltage equipment.

As shown in Fig. 8, the initial impulse distribution in a transformer winding can be obtained with the electrolytic tank, but it should be remembered that it is only correct for an infinitely steep wavefront that would charge all the winding capacitances before current flowed in the self-inductances, and it does not give information on the subsequent oscillations which must occur unless the initial distribution is uniform. These oscillations usually cause significant stresses further down the winding, whereas the field plot only shows stress concentrations at the beginning of the winding. However, the field plots enable the improvement in the initial distribution effected by different shield designs or winding dispositions to be easily determined. As improvement in the initial distribution results in reduction of the oscillations, I agree that useful information can be obtained in this way.

Mr. G. F. G. Clough (*at Manchester*): The problem of fluid flow differs from the problems represented by eqns. (1)–(4) of the paper in that the fluid possesses mass. It would appear, therefore, that the streamlines shown in Fig. 18 refer to the flow of a fluid which, in addition to being incompressible, possesses no mass. Can the authors indicate how this difficulty can be overcome in problems of fluid flow?

What standard of cleanliness is necessary for the metallic parts of the models in order to obtain consistent results?

Mr. N. Ashton (*at Manchester*): From a brief inspection of Fig. 8 it appears to me that these plots would not have been very different had only the 0 and 100% electrodes been used and the turns of the h.v. coils been omitted.

I am very interested in the plots of current-transformer leakage fields shown in Fig. 14, but these problems are very rarely 2-dimensional since bends and loops in the primary conductors introduce a 3-dimensional aspect. Also, these problems are often further complicated by several cores of different magnetic materials being mounted on the common primary conductor or coil. Do the authors know any method of representing, in the electrolytic tank, permeabilities ranging from unity to several thousand?

Prof. F. M. Bruce (*at Glasgow*): In colleges the tank method is usually used with a metal sheet of resistive material in place of the electrolyte—this is convenient and clean, and sufficiently flexible for carrying out measurements with simple field configurations that can then be checked by analysis. Thus the student is instructed in the fundamental theory and introduced to the technique; this is the correct order of emphasis for an undergraduate course. All forms of analogue require a sound knowledge of field theory before the technique can be applied to a given problem.

Could the authors indicate a suitable ratio between tank and model dimensions, and give any information on the possibility of representing an insulator with conducting glaze in tank tests? Have they any experience of using the tank for 3-dimensional models other than those which can be represented by a wedge tank?

In the application of the tank technique to fluid-flow problems, the restriction to non-compressible flow lines should have been given greater emphasis. The method is so attractive that it may be used under flow conditions to which it is not applicable, and in preference to other methods which are.

Mr. D. S. Gordon (*at Glasgow*): The electrolytic tank is a very versatile tool, and by elaboration, capable of considerable accuracy and the ability to solve automatically such problems as the determination of electron trajectories in complicated fields. However, for many purposes, perfectly adequate results can often be obtained from very crude apparatus.

For teaching, the use of liquids—particularly corrosive electrolytes—is to be avoided if possible, and where principles only are being illustrated, graphite-coated card yields good results and may well be preferred. The system may also be simplified by using a d.c. supply and a galvanometer as a detector. The graphite-coated paper has the further advantage that a second layer of narrow graphite strips sprayed on top of the base layer gives the effect of different permittivities at right-angles, without

the necessity for distorting the model. Simple techniques of this sort are sufficiently accurate for estimating seepage in soils where the constants can only be estimated approximately.

It would be useful if the authors, from their experience, would give some more practical information on electrolytes, materials used for boundaries and probes, the effect of meniscus and the type of detector employed. When it is required to obtain the greatest possible accuracy from a given size of tank, attention to details such as these is most important. This, of course, presupposes that deficiencies in the mechanical design of the tank and probe carriage do not limit the ultimate accuracy.

Mr. L. M. Haddow (*at Glasgow*): The authors mention the use of tap water. I have used the electrolytic tank for various simple field problems, and have found that tap water is not very satisfactory, since its very high resistance (which might be a feature of the Glasgow water supply used) makes detection more difficult. Copper electrodes with slightly-acid copper-sulphate solution were used, the supply frequency being 50 c/s. Polarization then caused some trouble, and I found that the probe voltage could not be completely balanced out by a sinusoidal voltage tapped off a resistance divider. There was always a complex residue which upset the detection. I also found that there could be quite considerable and uncertain voltage differences across the boundaries between the electrodes and the electrolyte. Have the authors encountered this difficulty?

Messrs. H. Diggle and **E. R. Hartill** (*in reply*): The object of the paper was to indicate in a general way some of the types of problems in electrical engineering which can be solved conveniently by the electrolytic bath. Many of these involve only 2-dimensional work, and considerations of space prevented our dealing with 3-dimensional technique. The lack of published information on deep-tank technique, to which Mr. Wainwright and Prof. Bruce refer, is largely due to the inherent difficulty in representing different permittivities. In the measurements on switchgear installations, to which Messrs. Burtt and Wainwright refer, the different permittivities have not been taken into account, so that the results can only serve as a guide.

Our work on the representation of different permittivities in 3-dimensional work is not sufficiently advanced for us to be able to report that we have reached a completely satisfactory solution, and even with the refinements we have in mind it is not possible to represent service conditions of pollution or high humidity.

In the deep tank which we have constructed, plots in a horizontal plane are made at any depth, using the type of equipment already described, together with a probe sealed into a 5 mm glass tube, from which it projects 2 mm. For plots in vertical planes the plotting table is swung into a vertical position immediately behind the probe. Vertical movement of the probe is achieved by a rack-and-pinion mechanism.

In reply to Mr. Gordon, we use brass electrodes coated with Aquadag; and steel needle probes have been found satisfactory for potential measurements to accuracies of 1 or 2%. For more stringent requirements in plotting gradients to $\pm 1\%$ accuracy or better we have sometimes used silver-plated electrodes and liquid probes* together with pulse excitation to reduce polarization effects. Tap water has been found suitable in either case, and we would expect the water used in the north-east by Mr. Burtt and in Glasgow by Mr. Haddon to give similar results.

We note Mr. Elliott's confirmation of concentric electrodes for calibration purposes, and his approval of the cross-wire method of plotting. Prof. Bradshaw's modification using a magnet system was described in a most interesting manner by Mr. Attree.

In reply to Mr. Shaw, we should perhaps have stated that the curvilinear square method was well established, and of more immediate usefulness perhaps than mathematical methods. His ingenious suggestion for 3-dimensional field plotting in colour sounds fascinating. Our own experience in this field has been limited to the recently developed sensitive crayons which we have used for surface temperature measurements. We would refer Mr. Shaw to a paper* which deals with temperature distributions in buildings and their assessment by means of the electrolytic tank, and Dr. Feinberg to McDonald's paper and to the contribution by Mr. Einstein regarding the simulation of dielectrics. The conducting paper mentioned by Dr. Feinberg suffers from non-uniformity.

We agree with Mr. Wright regarding initial impulse voltage distributions and on the effort now required to correlate field plots with breakdown voltages.

Messrs. Baker and Giles have pointed out some of the assumptions which must inevitably be made with an electrolytic tank model. Mr. Giles has also shown an interesting example of fluid flow through turbine blades in which circulation has been taken into account. The compressibility effects, mentioned by Prof. Bruce and Mr. Clough, can be accounted for,† but no method has yet been devised for simulating boundary-layer phenomena. Mr. Hodgetts has indicated errors involved in magnetic field plotting. The discrepancies in Fig. G were due to the inherently weak electrodynamic fields in the model and the difficulty of manoeuvring the plotting probe between the probes representing the winding. Increased accuracy could have been obtained by greater subdivision of the winding although this would entail more complicated support and additional resistors. Permeabilities ranging from 1 to 1 000 mentioned by Mr. Ashton can be simulated by using solutions of appropriate conductivity which may be contained in separate insulating cells having thin walls. The walls are provided with metallic studs or jumper strips which ensure contact between the main electrolyte and the solution in the cell. Mr. Burtt has also referred to the use of solid semi-conductors in this type of problem.

It is apparently impossible to represent in the electrolytic bath the 3-dimensional magnetic field of conductors which are not entirely rectilinear or completely circular. If such a representation were in fact feasible, the possibility would then arise of experimental verification of the Biot–Savart law for an element of circuit. It is perhaps significant, however, that the law is inherently incapable of experimental verification.‡

It is not possible to answer Prof. Bruce's query as to a suitable ratio between the tank and model dimensions, as it is dependent on the nature of the problem being studied; e.g. in Fig. 1 the boundary of the tank has no bearing on the conditions, whereas in Fig. 5 it is essential that it be at a considerable distance from the centre line of the bushing.

Mr. Haddon refers to polarization difficulties. These are overcome by the square-wave technique of Sander and Yates.§

* SANDER, K. F., and YATES, J. G.: "The Accurate Mapping of Electric Fields in an Electrolytic Tank," *Proceedings I.E.E.*, Paper No. 1370 M, August, 1952 (100, Part II, p. 167).

* "Study of Problems of Heat Transfer by Electrical Analysis," Building Research Congress, Division 3, 1951, pp. 1, 2, 3, 23 and 24.

† TAYLOR, G. I., and SHARMAN, C. F.: "A Mechanical Method for Solving Problems of Flow in Compressible Fluids," *Proceedings of the Royal Society, London*, A, 1928, **121**, p. 194.

‡ CARTER, G. W.: "The Electromagnetic Field in its Engineering Aspects" (Longmans), p. 110.

§ SANDER, K. F., and YATES, J. G.: "The Accurate Mapping of Electric Fields in an Electrolytic Tank," *Proceedings I.E.E.*, Paper No. 1370 M, August, 1952 (100, Part II, p. 167).

621.395.72.004.6

Paper No. 1777
Jan. 1955

MAINTENANCE PRINCIPLES FOR AUTOMATIC TELEPHONE EXCHANGE PLANT

By R. W. PALMER, Member.

(*The paper was first received* 16*th August, and in revised form* 12*th October*, 1954. *It was published in January*, 1955, *and was read before* THE INSTITUTION 10*th February*, 1955.)

SUMMARY

The maintenance of the complex mass of equipment which constitutes the modern automatic telephone exchange requires to be different from that of most engineering plant, mainly because the finding of the faults is so much more difficult than their subsequent repair. A comparison of maintenance practices throughout the world needs first to be qualified by differences in the design of systems and plant. With this reservation, the experiences in the British Post Office and abroad with overhauls, cleaning, routine testing, alarms and monitoring devices are then reviewed and compared. This leads to the compilation of a pattern of maintenance effort in relation to the effect on service and reveals the many alternative methods of prevention, disclosure and correction of service failures. The relative efficiency of methods of reducing the number of plant faults or reducing the number of calls affected by each plant fault is then discussed, and it is suggested that there may be a drift from the old-established policy of routine preventive maintenance in favour of a flexible and qualitative policy to suit the local condition of the plant. At the same time, the design of maintenance aids appropriate to modern technical developments in switching systems is reviewed. The paper concludes with a Bibliography and a list of definitions of terms that are used with a restricted meaning in the text.

(1) INTRODUCTION

Maintenance practice in automatic telephone exchanges can take many forms. To some specialists it means the design and provision of adequate testing equipment; to others it means adjustment instructions and techniques; there are those who concentrate on periodical overhauls and routine tests as the essence of preventive maintenance, while others stress the procedure for prompt handling of reported failures. The object of the paper is to make a critical review of these maintenance details in turn and to deduce some general principles from practical experience in the United Kingdom and abroad. It is not to be expected that a precise policy will be advocated for any one telephone system, but it is necessary to bring all the main factors into proper perspective so that the most efficient maintenance policy can be decided for any type of exchange plant and for any condition of service. The possible misunderstanding of commonly used maintenance terms is avoided by using the more precise definitions quoted in Section 10.

(1.1) Scope of Review

To keep this paper to readable proportions, it is not possible to give detailed technical descriptions of all the many maintenance facilities quoted; there is normally no difficulty in getting such information from technical journals and from the maintenance instructions published by telephone administrations. It has also been necessary to ignore many basic maintenance aids that obviously must be made available irrespective of controversial policies, but it is as well to mention a few of them by name. The maintenance staff must be provided with adequate and simple descriptions of the plant on which they are working (the British detached-contact circuit diagrams are an excellent example of this). Similarly, the methods of adjustment must be defined by instructions, the appropriate tools and spare parts must be made available and the staff must be trained in their use. Standard procedures, schedules and forms must be provided by every large-scale organization for the adequate control of the work. Finally, there are the many problems of organization, personnel and staffing which are outside the scope of an Institution paper.

(1.2) Maintenance Standards

By way of introduction, reference must also be made to the comparisons often made between the maintenance of telephone exchanges and that of other kinds of engineering plant.[1,2] The quality or performance to be maintained depends on the technical characteristics of the plant and on the service to be provided. It is obvious, for example, that the standard of performance of a single electric lamp in a key position should be much higher than for a battery of 100 floodlights in which an average of 1% failure is of little consequence. Again, it is a fallacy to assume that routine preventive maintenance in the accepted sense is always practical or even effective. Continuing the analogy of illumination, if the daily testing and cleaning appropriate to an oil lamp were applied to an incandescent-filament lamp, the number of failures in service would be greater, not less. With these factors in mind, any review of telephone exchange maintenance must commence with a study of variations in plant design and function. Moreover, the design and manufacture of apparatus and of systems can also materially contribute to maintenance efficiency if the requirements are fully understood.

(2) PLANT DESIGN

(2.1) Exchange Design

It is generally accepted that fault location in automatic telephone equipment is easiest when each stage of the setting-up of calls digit by digit is a self-contained operation, such as in the non-director step-by-step systems [Fig. 1(*a*)]. By contrast, the designs that incorporate directors, registers, translators, markers or other controlling equipments that are dropped out of circuit as soon as the connection has been set up [Fig. 1(*b*)] introduce special maintenance problems in varying degree.

For economy in the amount of switching plant to be provided, it is also essential to design the exchange trunking scheme on the basis of graded outlets, but striving for the last ounce of trunking efficiency is accompanied by serious disadvantages in service and in maintenance cost. It is significant, for example, that the British Post Office has decided to forgo the ideal grading of uniselector outlets in favour of having two "home" positions, thereby avoiding at little cost the severe service handicap when a first-choice outlet is faulty during periods of light traffic. Even so, all gradings are designed to provide a given grade of service during the heaviest busy hour, not for average or continuous conditions, and the resultant wide variations in traffic per switch disturb the economics of standardized maintenance attention. The traffic distribution in a typical group-selector grading is shown in Fig. 2. The curve shows the distribution of busy-hour traffic (calculated traffic units) and it is seen that the load is fairly evenly shared by three-quarters of the total plant. The pillar-

Mr. Palmer is in the Post Office Engineering Department.

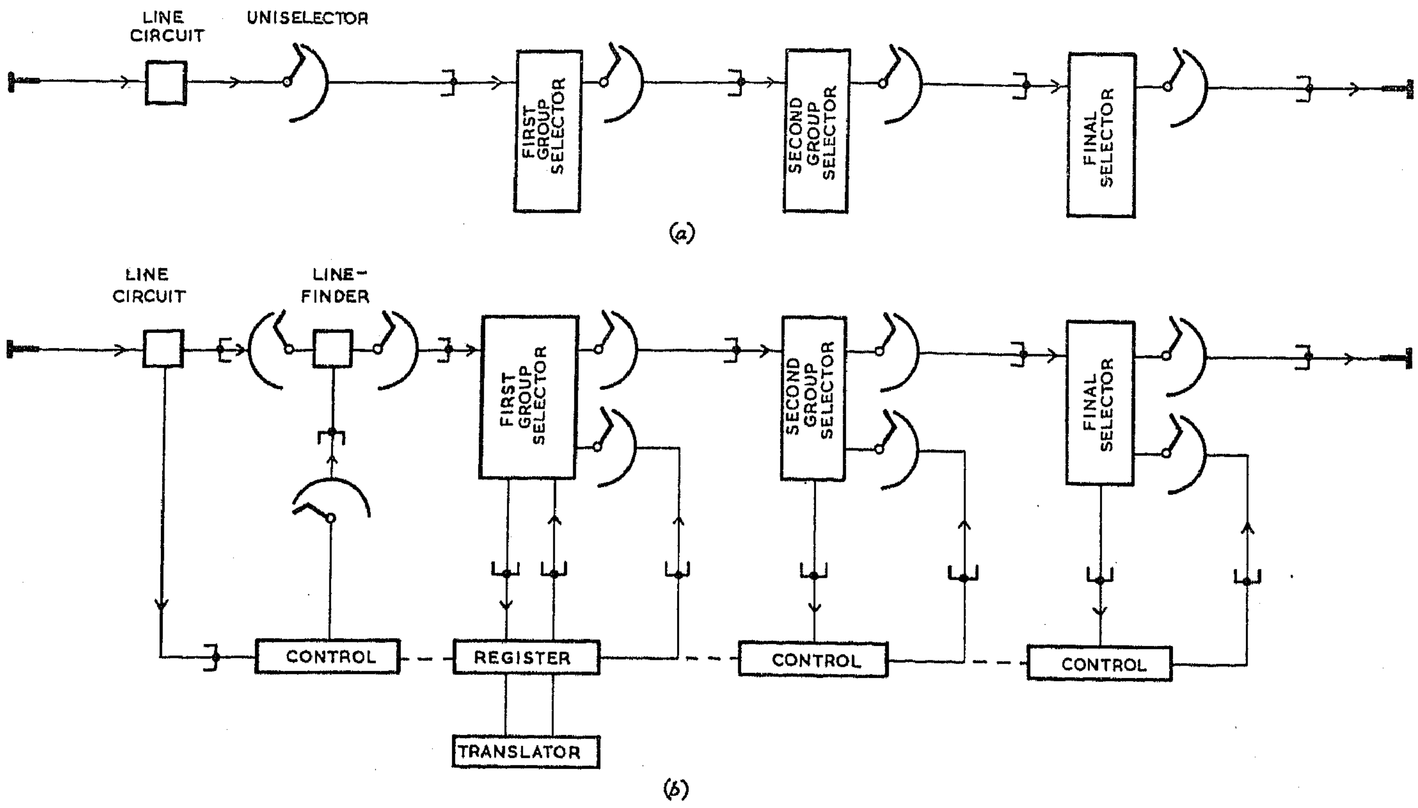


Fig. 1.—Complexity of typical trunking.

(a) Step-by-step system.
(b) Common-control system.

graphs show the typical weekly traffic (measured number of operations). From this it is deduced that the first 25% of selectors may operate 60 000 times in the course of a year while the last 25% are relatively idle at an average of 1 000 calls a year. Thus all faults do not have the same effect on service. If the economic frequency of routine maintenance is established for average traffic and average fault liability, it is bound to incur some waste of effort on the late choices and to be inadequate for the early choices in the gradings.

(2.2) Apparatus Design

Apart from the ways in which selecting devices may be arranged and used in the exchange trunking scheme, there are wide variations in types of switching mechanism. Among those that are primarily mechanical in type, different maintenance techniques are required for step-by-step mechanisms (Fig. 3), power-driven mechanisms (Fig. 4), and crossbar selectors (Fig. 5). There will be even greater differences in the maintenance requirements of electronic devices.

Strength, reliability and stability are obvious requirements, although they must be related to first cost. Mechanical selectors have an almost indefinite life if piece-parts are replaced individually as they wear out, but a figure of two million calls is a typical life of an ordinary step-by-step group selector in practice. The infrequency of failure is more important than ultimate life, and any specified life tests should carefully define what routine readjustments and other preventive maintenance are required or are to be permitted. It is also necessary to define what constitutes a service failure (irrespective of wear) such as will justify renewal of component parts or other interference in the course of the life test.

Accessibility for readjustment or replacement is not necessarily an important maintenance advantage; it depends on the fault liability. At its worst, maintenance accessibility may provide a temptation to interfere with something that is best left alone, but if some regular maintenance attention is essential—such as lubrication, cleaning or readjustment—then accessibility becomes an important feature in design.

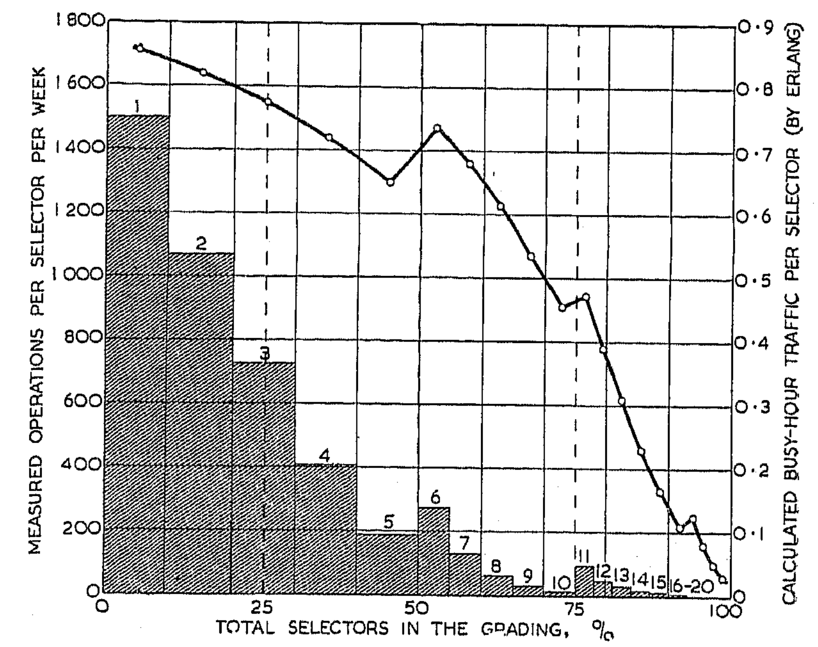


Fig. 2.—Typical distribution of traffic in a grading.

Widths of pillargraphs represent proportions of first choice, second choice, etc., in the total quantity of selectors provided. This example refers to a 6-group 20-outlet grading of second-group selectors at a non-director exchange.
Number of trunks in the grading = 61.
Busy-hour traffic capacity of the grading = 36 traffic units.

(2.3) Circuit Design and Tolerances

The reliability and need for maintenance attention of any exchange apparatus cannot be fully defined or judged except in relation to the demands made on it by the circuits. This is largely a matter of tolerances on current adjustment and timing, and it accounts for a great deal of apparent conflict in reports on mechanical performance, particularly in comparisons between laboratory tests and field performance. In one sense there is no such thing as a "circuit fault" or an "electrical fault," for all are caused by some mechanical change of condition. In another sense, there are many reported mechanical "failures" of dimension or adjustment that ought to be classed as circuit faults because of unreasonable circuit requirements. Reduced factors of safety either increase the frequency of failure in service or

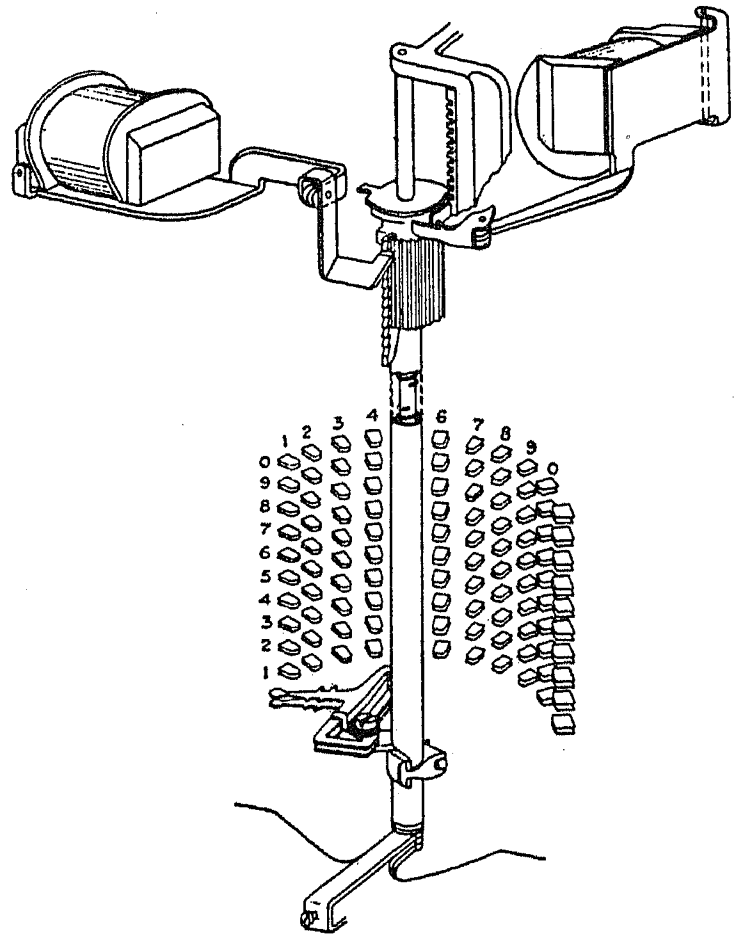


Fig. 3.—Basic ratchet-and-pawl selector mechanism.

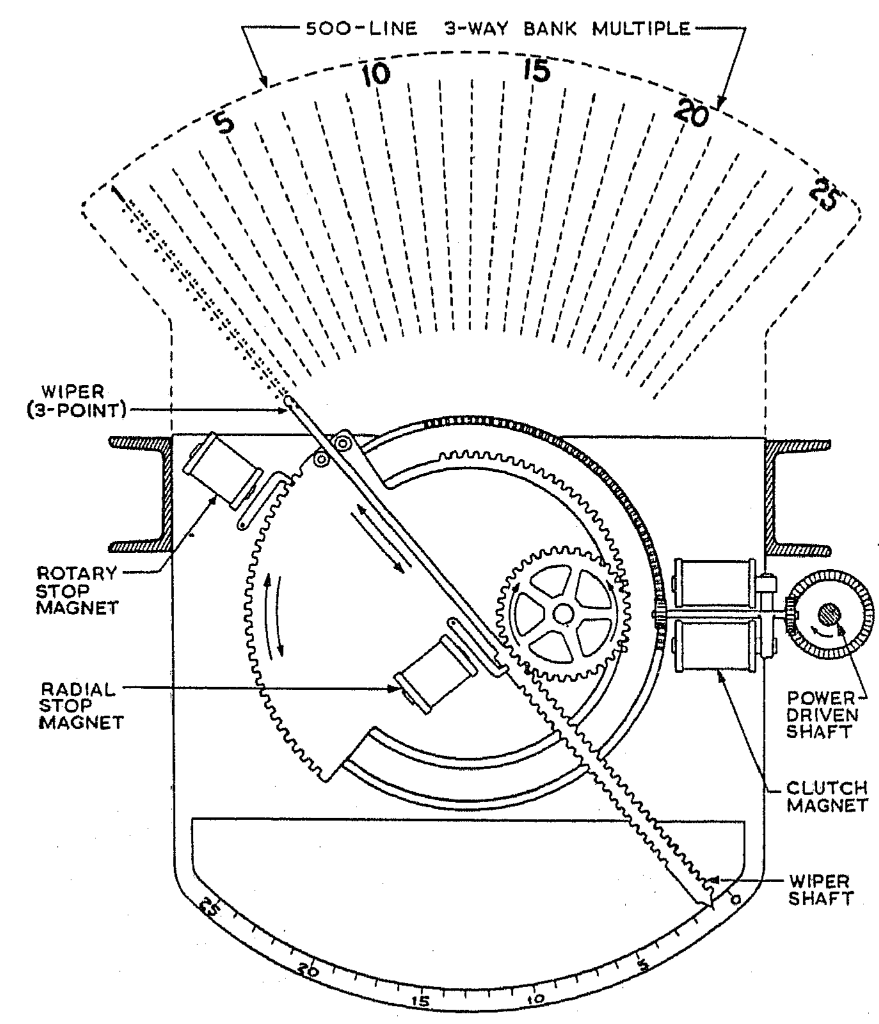


Fig. 4.—500-line power-driven selector mechanism.

lead to increased maintenance expenditure in the form of more frequent overhauls, more stringent tests and the replacement of parts when only part-worn.

The determination of acceptance-test standards calls for an economic balance between the cost of manufacture and the expected life of the plant before maintenance attention will be needed. The maintenance aim will naturally be to make the acceptance tolerances as tight as practicable, irrespective of the extent to which they can be allowed to deteriorate before the apparatus fails in service. If, then, a particular circuit permits the standard mechanism to function correctly even when well worn [Fig. 6(*a*)], the test tolerances used for maintenance ought to be much wider than the acceptance limit, and so avoid unnecessary readjustment. The more severe the circuit applications, the closer will be the maintenance tolerances on that same mechanism [Fig. 6(*b*)], leading to more-frequent readjustment. This is a statement of the maintenance ideal, whereas it is more usual to make acceptance and maintenance test limits the same; but the principle is stated, since it has a bearing on the use of such tolerances in overhauls and other maintenance work.

(3) OVERHAULS AND CONDITION OF PLANT

(3.1) Experience with Overhauls

The service given by a telephone exchange is dependent to a large extent on the physical condition of its apparatus. This seems to imply the need for inspections and overhauls at regular intervals, and this kind of routine preventive maintenance has been conducted in most countries for many years. There have been remarkably successful results in some British telephone exchanges recently with concentration on the quality of mechanical overhauls above all other maintenance operations. To make this scheme of maintenance possible without increasing the staff, the functional routine testing was either abandoned or drastically reduced. This is a separate point, to be discussed later, but the immediate concern is that fault liability has been substantially reduced, particularly in unattended rural equipments. Then, as the second and subsequent cycles of overhaul required less attention, it has gradually been found possible to reduce the frequency of overhauls and to reduce the man-power substantially, while still achieving a better quality of service than before. Economies have also been made by relating the frequency of overhaul of particular switches to the traffic carried (i.e. position in a grading), or to a fault history associated with each switch.

(3.2) Experience without Overhauls

In complete opposition to the above experience it is noted that the Bell System in America have ceased regular overhauls of step-by-step mechanisms. They have operated a rather elaborate fault-analysis scheme to bring defects to notice, but the maintenance man is not expected to do any more than is necessary to clear each observed defect (actual or suspected), leaving everything else undisturbed.

A similar outlook is observed in Sweden for their 500-line power-driven selectors. A so-called "overhaul" is given every few years, but it consists almost entirely of meticulous cleaning without readjustment, except at rural exchanges. Each mechanism is washed with a cleaning fluid and appears immaculate and highly polished. A superficial mechanical inspection is then made, and the only adjustment check is a jig measurement to ensure proper engagement with the power drives on the racks; this is mainly a measure of the wear of bearings. This concentration on extreme cleanliness is not practised anywhere else to the same extent.

Some laboratory experiments have been made by a British manufacturer, in an attempt to simulate maintenance conditions, with and without overhauls, and the results have been published in some detail.[16] Some step-by-step group selectors were first given a carefully controlled life-test without maintenance readjustment or overhauls other than lubrication and

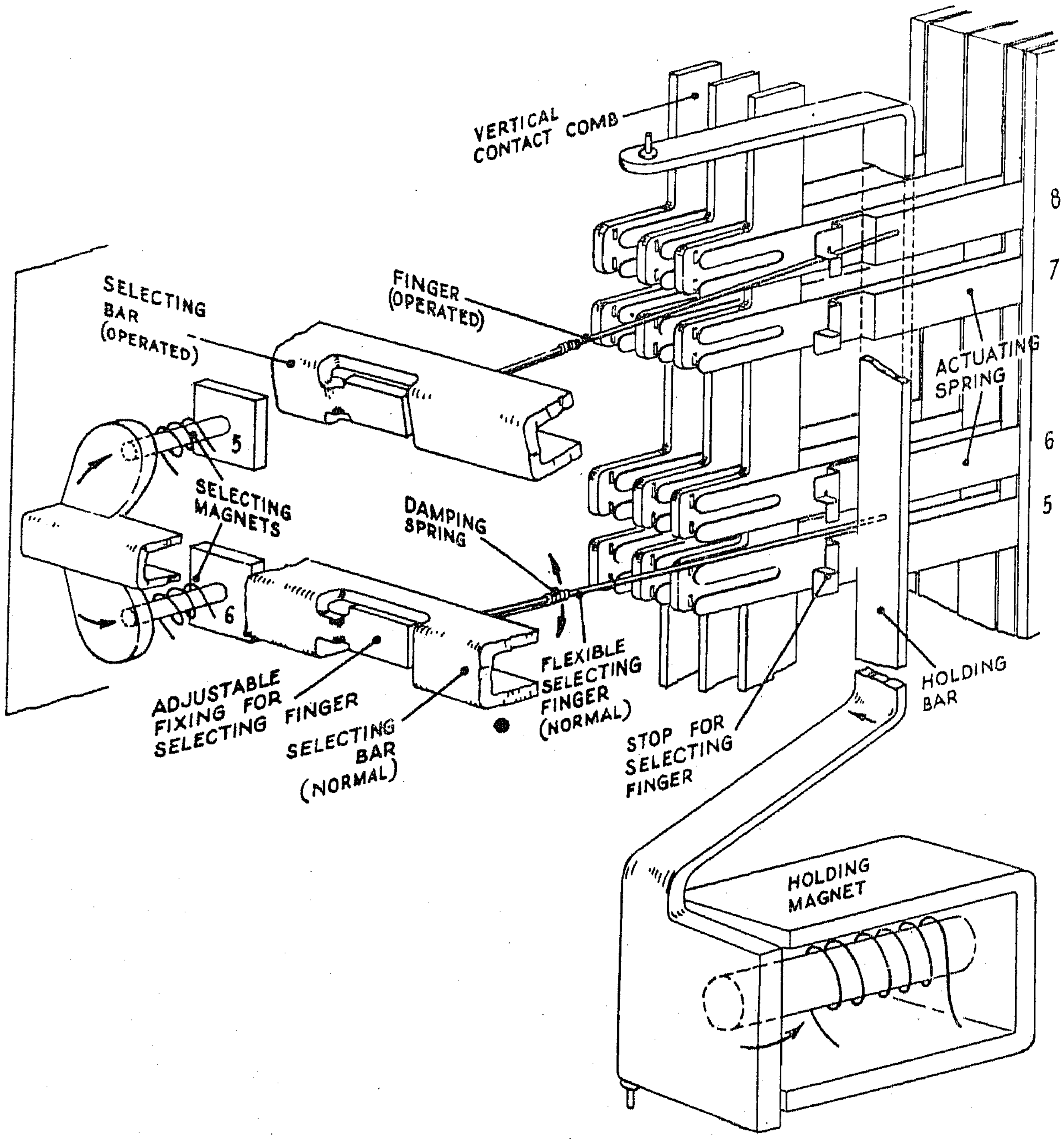


Fig. 5.—Typical crossbar switch mechanism.

bank cleaning. At the end of two million operations—equivalent to 40 years' life under average exchange conditions—there had been one service failure per selector every $4\frac{1}{2}$ years. The complete test was then repeated with a fresh set of selectors, but this time with overhauls and readjustment at the equivalent of yearly intervals. Every defect found during overhaul or inspection was listed as a routine plant fault, whereas stoppages of the test cycle were listed separately as service failures. The number of service failures was not reduced, but was $2\frac{1}{2}$ times as great as when no preventive maintenance was applied. The inference was that the overhauls caused more failures than they prevented. Accelerated laboratory tests cannot simulate exchange conditions entirely, but they indicate that the bearing and impacting surfaces will "bed in" if a selector mechanism can be allowed to remain undisturbed, and the switch is then capable of many years' life without overhaul.

(3.3) Condition of Plant

To resolve the apparent contradictions in the above evidence, one must first make a clear distinction between routine overhauls that are repeated at short intervals and those that are special and are made as required. Experience confirms very clearly that good service is dependent on good mechanical condition, irrespective of what testing and other maintenance work is performed. It follows that the need for an initial overhaul (and its thoroughness) depends on the state of the plant at the

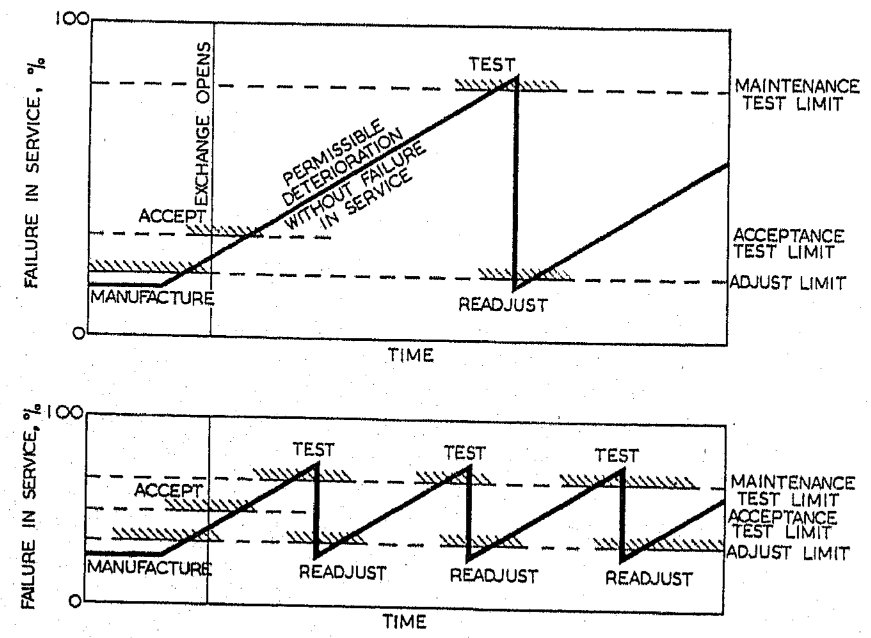


Fig. 6.—Comparison of adjustment tolerances.

(a) Wide maintenance-test tolerances for a component when used in an easy circuit application.

(b) Narrow maintenance-test tolerance for the same component when used in a difficult circuit application.

time, being influenced by the condition of adjustments when originally installed and by subsequent good or bad maintenance. If the plant is not already in good adjustment appropriate to the

circuit demands made on it, a single and thorough overhaul is necessary to make it so. Thereafter, the plant will give the best result if left alone so long as it is giving satisfactory service. There may, however, be a special case for regular but infrequent overhauls in rural areas if a single plant fault at a remote unattended exchange may cause a major breakdown of the exchange service; the cost of special attendance of an expert to correct each individual fault as it occurs is usually excessive.

(4) DUST AND VENTILATION

(4.1) The Effects of Dust

Apart from extremes of temperature and humidity, the greatest effect of atmospheric conditions on maintenance reliability arises from dust. It accounts for a large number of faults described as "dirty contacts" on relays, wipers and similar connections; it also causes abnormal wear of sliding or impacting parts, and it may contribute to the sticking (failure to release) of relay armatures. Careful analysis of the dust in typical telephone exchanges shows that it varies from long fibres visible to the naked eye, through a range of "cinder" and other solid material, down to particles of soot of microscopic dimensions. It is the gritty dust that increases the rate of wear of sliding surfaces, especially if combined with oil or grease, but the false disconnection of light-duty relay contacts is largely due to lint and other fibrous dust larger than 25 microns in size.[1]

(4.2) Ventilation Systems

Most large exchanges are provided with air-filtration plant because they are usually situated in towns where atmospheric dust is considerable. A costly installation of electrostatic precipitation is rarely justified, and it is usual to provide either screen-type, or the even cheaper impingement-type, filters. The efficiency of air filters for telephone exchanges can be judged only in relation to the kinds of dust that cause apparatus faults, and ventilating systems should not be condemned merely because test cloths across the outlets become black with smoke and fine soot. Particles of carbon less than 2 microns in size have a negligible effect on the performance of telephone-exchange plant.

(4.3) Local Generation of Dust

Even perfect air filtration is of little value if dangerous dust is generated within the exchange apparatus room. It is most important to use wires and cables that do not shed lint and dust whenever they are handled. All kinds of apparatus and wiring should be either capable of being cleaned efficiently or should be enclosed so that the dust deposits cannot be scattered. Desks, filing cabinets, clothes lockers and other dust-producing items should be kept out of the apparatus rooms. Main distribution frames, where dust is being disturbed by daily rearrangements, should also be in a separate room.

(4.4) Cleaning and Protection

However good the precautions may be, there will still appear some dust and dirt on the apparatus and wiring. All cleaning should be infrequent, thorough and self-contained. Casual wielding of a brush or dry duster will lift much of the dust into the air, only to fall elsewhere, and the cleaning cloths may even shed dust and lint on the apparatus that is being "cleaned." Cleaning with an oil or solvent is the appropriate method for most smooth surfaces, and is preferable to ordinary vacuum sweeping. For general cleaning of racks, cables and wiring, the most satisfactory process is the use of a compressed-air gun in conjunction with adequate suction and shrouds. If this is not practicable, it may be better to leave the dust undisturbed. The provision of easily-cleaned covers will limit the disturbance and deposition of dust on vital components, but it is rarely necessary, or even desirable, for the covers to be scientifically dust-tight. Thermal baffles within the cover are particularly effective in reducing dust circulation.

(5) FUNCTIONAL TESTING AND ALARMS

(5.1) Functional Testers and Routiners

In the course of development of routine testing techniques for automatic switching apparatus, the original simple test-boxes have grown in most countries into large semi-automatic testers mounted on wheels. Testing equipments for the larger and more complex exchanges have been developed still further and have become rack-mounted "routiners" with automatic access equipment. In spite of high capital cost, it has been established without question that automatic routiners are economically justified as compared with manual testing if there is a large amount of switching plant and if the testing is required to be performed at very frequent intervals. Where these conditions have applied, notably for the British director system, automatic routiners have been widely used.[17] A modern development in routiner policy as applied to very large exchanges is the addition of a fault recorder for printing the dockets automatically.[19] By the use of this equipment, which can also start and stop the routiners at preset times, the staff are not interrupted every time the tester encounters a fault, and testing can proceed without delay even when there is no staff in attendance.

(5.2) Routine Test Frequencies

Since the early days of automatic telephony, almost all telephone administrations have produced schedules of recommended routine test frequencies for the various items of plant, the frequencies representing a balance between the cost of the maintenance work and the desire to find faults before they can seriously affect the quality of service to the public. The frequency of testing can often be reduced to suit local requirements, so it is most important that the maintenance organization and procedure shall not be so rigid as to cause the staff to play for safety and keep to nominal schedules. Functional testing, by itself, does not improve the physical condition of the plant (it can only indicate where plant faults exist), so an increase in test frequency at the bad exchange may reduce the time available for careful prevention and correction of the plant faults already known. Contrary to the general trend towards less frequent use of manual test-boxes, the use of automatic routiners is usually raised to the practical limit, since faults will then be found with a minimum of delay and with no appreciable increase in man-power. Daily (or nightly) tests of common equipment such as the director train of switches is quite normal. Under such conditions a critical factor is the number of times the plant has to be corrected to pass the routiner, and this is related to the number of operational features tested (including the minor and less used facilities) and also the stringency of the tests.

(5.3) Stringent Testing

The only kind of testing that "prevents" service failures is that which applies some form of marginal testing more stringent than normal service conditions, but this can only cover the failures which are preceded by some progressive and measurable deterioration. Unpredictable failures such as contact faults, disconnected wires or broken parts cannot be prevented by any functional test, and not always by inspection or overhaul. In practice, the bulk of the tests applied can have no marginal limits (the switch either works or it does not), whereas the main application of

marginal tests is in the pulsing and stepping; it is relatively easy to provide stringent test conditions by variation of line characteristics and of pulsing speed and ratio, but marginal tests of current and timing performance of components within a selector are expensive, in terms of both test gear and access leads.

(5.4) Route Testers

The separate test boxes or rack-mounted routiners so far considered do not normally check the condition of the outlets and wiring between one rank of selectors and the next, nor do they test units of apparatus in combination, so there has been developed in the British Post Office what are called "artificial traffic equipments"[6] that can be connected to selected subscribers' exchange-line equipments and will automatically originate a rapid sequence of calls from any one line to any other. There is also a facility such that calls can be directed to special answering units on test lines at a distant exchange. If any test call fails to reach its intended destination, the test cycle is stopped and an alarm condition is established; this covers the main functions of individual testers or routiners plus the testing of cabling and wiring. A better title for such equipment would therefore be "automatic system tester" or "route tester." Similar testers are used in Sweden,[8] but are connected only to one pair of lines and repeat the calling train of pulses again and again. The so-called "test frames" in some early crossbar exchanges in the United States[7] also comprise end-to-end route testers with functions similar to those in Sweden.

There are three ways in which automatic route testers can be used. They can be employed for routine fault-finding, with perhaps some manual testers for supplementary diagnosis on individual switches. Alternatively, the route testers can be used with great benefit as an occasional supplement to the ordinary routine testing. The third use is for observation only, the failures being noted but not traced; this is a duplication of the main functions of service observations that are made on live traffic by telephonists, but local engineering staff may prefer the observations by this automatic device, because a larger sample of calls can be made at less cost and the results are immediately available for special action if abnormal trouble is revealed.

(5.5) Exchange Alarms

It is often forgotten that the application of routine test conditions to the plant is not the only kind of functional testing that can be performed. There are many schemes of "self-testing" that apply continuous checks, such facilities being built-in to the design of the system and giving an immediate alarm on the first occasion of failure. To a very limited degree this facility is provided in British telephone equipment by the scheme of exchange alarms for various failures of common services, including power failure, ringing failure and the operation of distribution fuses. The alarm is also sounded if a called subscriber is held by the caller failing to clear, and if more than a certain proportion of first selectors are held by permanent calling signals (probably by line faults). A similar range of fault conditions is tested in the United States, but their subdivision into about five degrees of urgency[1] has much to commend it, particularly for the alarms extended from distant unattended exchanges.

(5.6) Self-testing Circuits

There are many switching circuits that are self-checking in the sense that they will cause traffic to by-pass faulty outlets or will provide an alternative path when certain fault conditions are encountered. The "battery-testing" hunting circuit is a simple illustration of this, and "revertive" impulsing is an outstanding example of the self-checking principle applied to a complete switching system. Such facilities give an improved service to the subscriber but require supplementary tests or observation for the plant faults to be detected and located. There is also a self-checking facility in the British director system, where failure to complete its operations causes the director to be released and the "number unobtainable" tone to be sent back to the caller.

In some other systems, notably Panel and 7A Rotary, a self-indicating facility is added. On the failure of any item of common equipment (or the detection of failure of the distant equipment controlled by it), the register is held and prevented from handling further calls, and an alarm is given by a bell or a lamp; meanwhile, the calling subscriber is released. The setting of the register at the time of failure can then be studied, even if the route that had been set up has been released.

(5.7) Manual Monitoring Devices

As an alternative to the holding of the register on the occasion of every failure, all the common controlling equipments may be connected to lamps on a central panel, so that a monitoring officer can use discretion in holding a proportion of the failures for investigation. In the Swedish 500-line system, for example, the lamps indicate seizure and various operations of the register, and one of them—a red lamp—indicates delayed sending of the translation. On examination of the panel it is therefore easy to pick out the glowing red lamps and to notice any one that remains alight for more than a few seconds. By depression of the associated key the route can be held against release and a central panel displays the numbers so far dialled into that register and the numbers so far dialled out. If the delay is caused by the subscriber, it is possible to hold the key depressed and to watch any delayed digits coming in, one by one, and to see the translation go out accordingly. In case of subscribers' error or other query, the observer can plug in to the circuit and speak, advising the caller and assisting him in his dialling, or checking the actions of the subscriber against the performance of the register. If then a plant fault is indicated or suspected, the route set up can be locked permanently for tracing, while the caller is immediately freed to make a fresh call.

(5.8) Automatic Monitoring Devices

The efficiency of visual monitoring depends very much on the time devoted to watching the pattern of lamp signals and on the skill of the observer. Nevertheless, this kind of fault location has been so successful in Sweden and in the United States that both countries have developed automatic monitoring devices. In Swedish A204 crossbar exchanges,[13] all service failures at the common control points are noted on a "centralograph"—a recording instrument capable of marking rows of dots on a continuous roll of paper. If a fault occurs, such as failure to receive a "proceed to send" signal from the next stage in the network, the reference number and the impulse setting of the common control equipment are recorded in code and the whole connection is then cleared down by forced release. The reliability of the switching system is such that in normal times only about three faults per week are recorded in a large exchange. In practice, weeks may go by without a single fault record, and then a batch of them will suddenly appear. It is for this reason that the centralograph is normally unattended except for casual inspection, but an audible alarm is given if 25 successive faults are recorded. Such an event requires prompt attention, and the data from 25 faults generally lead to immediate localization.

The equivalent "automatic monitor" introduced with the American Crossbar No. 5 system provides fault data on punched cards.[5,12] It is also different in that only a percentage of calls is monitored. As each sender or register is taken into use it con-

nects itself to the common monitor only when the monitor happens to be free. It is stated that about 17% of calls are so observed. The monitoring device is used continuously, whether the exchange is staffed or not, and the recorded results are analysed in the same manner as in Sweden. There is also associated a master test frame[18] at which artificial traffic can be originated on any route and the results recorded on the automatic monitoring equipment.

(6) MAINTENANCE CONTROLS

(6.1) Fault Complaint Service

In the maintenance of automatic switching equipment it must be accepted that the reporting of troubles by subscribers is of very little help. The customer rarely makes specific complaints of calls that he has to dial more than once, and it is very difficult to locate faults in the middle of the switching network even when he does. Almost all of the complaints received are about complete loss of service or troubles confined to the subscriber's line or instrument. On the other hand, a well-conducted complaint service can supply general data on the overall quality of exchange service, and a proportion of faults can be found by tracing faulty calls through exchanges of the simpler types. In this connection it should be noted that there is normally compiled a statistic called "faults per telephone per annum" which could be more correctly entitled "service complaints per telephone per annum." All plant faults discovered by the engineering staff are excluded, but failures reported by telephonists are considered as "complaints" in some countries and not in others.

(6.2) Exchange Fault Control

The business of fault control within the switching plant must therefore depend mainly on a local engineering organization based on independent observation of the state of the plant and of the quality of service being given. Whatever machinery may be available in the form of testers or statistical data, the maintenance men must be technically qualified and much of their work will depend on local initiative and understanding of service needs. Beyond that, sufficient has already been said to demonstrate the need to watch economics as well as the quality of the telephone service.

(6.3) Service Observations

The most obvious and the most common method of determining the quality of service being given is by service observations—listening by telephonists on a proportion of the calls as they pass through the exchange. From these observations is derived a "percentage of calls lost due to plant faults," but the resultant figures published by various telephone administrations against this or equivalent titles do not always have the same meaning or accuracy. It may be considered uneconomic to observe other than on business lines and in busy areas. It has already been seen from Fig. 2 that the plant is used in different proportions in the busy hour as compared with slack periods. Not all of the call failures may be trapped if the observation is made at some intermediate point in the network instead of direct on the subscribers' lines, and additional faults may be created by the observation equipment tied to the circuits. Even the calculations can be distorted slightly by basing the percentages on the number of calls irrespective of the number of attempts per call, and some "faults" are much more serious than others.[3,15] Finally, the results may be distorted if faults are cleared on the observed traffic more promptly than under normal maintenance conditions.

It should be realized that, although accurate service observations may be useful as an overall performance statistic when taken over an extended period, the percentage of calls lost will fluctuate violently from day to day. In a well-maintained exchange there should be some days when the number of undetected plant faults is nil, and then a single disconnected wire on a first-choice switch or route may cause hundreds of lost calls until it is discovered and thrown out of service; the percentage lost calls will suddenly rise and will equally suddenly drop to zero when the plant fault is corrected. These wild fluctuations are far more significant for the local maintenance staff than for administrative control. For such reasons it is understandable that many engineers prefer testing with artificial traffic or the provision of automatic monitoring devices, because these reveal more promptly the plant faults awaiting disclosure and correction.

(6.4) Plant-Fault Statistics

An entirely different record is that of the plant faults that have been found and corrected, whether as a result of reports of service failure or in the course of preventive maintenance, but even this statistic must be treated with caution. If a switch is under repair for a specific failure, and the opportunity is taken to recondition three other details in urgent need of attention, does that represent one "plant fault" or four? Whatever may be the regulation answer to such a question, it is apparent that the fault statistics are merely aids to local appreciation of plant condition. Analysis of types of fault will indicate where maintenance effort should be intensified or could be relaxed, and the study of repeat faults on individual switches or ranks will provide guidance on the effectiveness of previous maintenance attention.

(6.5) Qualitative Analysis

The modification of schedules of routine maintenance in terms of local fault analysis is a trend observed in all countries, including the United Kingdom, but a large-scale effort has recently been made by the Bell System in America and has been given the name "qualitative maintenance."[1] It has been applied to step-by-step, panel and crossbar exchanges. Almost all regular functional testing and routine overhauling of apparatus is stopped (at least temporarily), and the local staff are required to analyse the statistics and other indications to justify almost every maintenance function that they undertake. If the pattern of service observations, service complaints and exchange alarms indicates the need for attention to some particular section of the plant, a snap check of adjustments or a local functional test is initiated and followed by a limited overhaul of that plant and on that particular occasion. It is then possible to see from the statistics the improvement achieved, and where next the efforts of the staff should be directed. After a year or two of careful observation it is generally found that the local staff have re-instituted some of the routine functional testing, but at relatively long intervals and at a varying intensity according to local requirements. Most of the testing that is still scheduled is specified to be "as required," and the remainder is rarely done more frequently than "yearly." Provided that overall results are satisfactory there is no cause for a controlling officer to criticize any exchange merely because its frequency of routines is high or low compared with others.

(6.6) Cost Control

There is no simple method of comparing maintenance costs at different exchanges. The cost per subscriber's line will depend on the number of switches per call, on the traffic density, on the type of plant and on many miscellaneous factors. The maintenance man-power on systems throughout the world has been quoted as varying from 0·5 to 5·0 men per 1 000 lines,[4,14] but the technical staff employed directly on the local switching equipment generally tends towards the lower figure. Such

variations in maintenance costs should be apparent from a comparison of tariffs in different countries,[11] but the comparisons are distorted by arbitrary rates of exchange of currency.

For domestic purposes it is more usual to measure maintenance-cost trends in man-hours per "work unit." The work unit allotted for each item of plant represents an arbitrary labour rating calculated or estimated to be required for its maintenance, but the aggregate for any complete exchange is not intended to be used for deciding the number of maintenance staff to be employed. The ratio of actual man-hours per work unit, however, gives a rough indication of where maintenance efficiency (high or low) might be worthy of investigation, and it certainly provides an accurate measurement of improved efficiency at any one place in comparison with past performance. The scope for reducing costs measured on this basis is illustrated by many countries achieving savings of as much as 50% over a period of a few years as a result of changes in maintenance policy. Whether such economies could be made everywhere depends on the efficiency that exists already.

(7) GENERAL PRINCIPLES

(7.1) The Pattern of Maintenance Effort

The first principle deduced from this review of maintenance operations is that in automatic telephone exchange practice it is not enough merely to have a low numerical fault liability of individual items of apparatus. Circuit requirements and the effect on service are more important than the physical fault. Moreover, to judge the efficiency of routine testing, of overhauls, or of any other maintenance procedure as a separate entity is misleading, if not impossible. It is only the interwoven pattern of design and maintenance effort which can be said to be efficient or inefficient. In Fig. 7 an attempt is made to classify the many alternative details of maintenance work in terms of their contribution towards service quality, and it will be seen that there are three main objectives, namely

Prevention of service failure (including prevention of plant faults).
Disclosure and location of service failure.
Correction of service failure (and of plant faults).

There are two decisions to be made: the proportions of effort to be devoted to these three functions, and the choice of method to be adopted for each separate function; the two decisions are obviously interdependent.

If a reasonable service at reasonable cost is the aim, it is necessary to consider to what extent prompt disclosure and correction may be cheaper and more effective than prevention of faults. It will depend on the maintenance facilities provided and on the relative ease of performing the different maintenance functions in any particular telephone system. These factors are now reviewed in the following Sections, bearing in mind that the improvement in service quality is just the same when the number of plant faults is halved as when the number of calls lost by each plant fault is reduced by 50%.

(7.2) Prevention of Service Failure

Prevention is better than cure when perfection is desirable and possible, but such an ideal in commercial practice must be modified by the economic factor. When deciding how far preventive maintenance in telephone exchanges is justified, three alternative ways of achieving the same object must be considered: improving the inherent reliability of the plant by technical design; anticipating or preventing deterioration by overhauls, lubrication and cleaning; obtaining warning of defects before actual failure occurs, which can be had by periodic tests, inspections and observations but is limited to the kind of fault that is preceded by some measurable deterioration. The strongest case for concentration on preventive maintenance for telephone-exchange plant (or for a particular portion of that plant) is when

(a) Failure in service is serious or costly.
(b) Detection of service failure is difficult or slow.
(c) Deterioration of plant is progressive and predictable.
(d) Additional faults will not be caused by the preventive operations.

In general, the trend in plant design is towards greater inherent reliability and a smaller proportion of predictable faults. For this reason alone the measures to prevent faults are becoming less effective, and it may be preferable to leave the plant undisturbed until failure occurs in service.

(7.3) Correction of Service Failure

The opposite of preventive maintenance is often said to be corrective maintenance, but such a statement is not always true for telephone switching plant. A proportion of the corrective work is essentially a sequel to preventive maintenance (this relationship is omitted from Fig. 7 for the sake of clarity).

Whether or not the corrective maintenance follows preventive procedures or actual service failure, it is usually just as effective (temporarily) to provide prompt alternative routing as to repair the specific plant fault. That is why the various ways of alleviating the effect of faults are shown as a separate group in the pattern of maintenance effort, and it follows that speed of repair may be less important than skill and accuracy in execution. The facilities and skill employed under the heading of corrective maintenance can also contribute indirectly to the prevention of service failure.

(7.4) Disclosure and Location of Service Failure

In telephone-exchange practice corrective maintenance requires special procedures to detect whether troubles exist and, if so, where. As the size or complexity of the system grows, particularly with the introduction of common-control equipment, so the plant faults that slip through the preventive-maintenance net may increase in number. The effect of each fault on the quality of service is also increased, and the prompt disclosure and location of service failures becomes increasingly important. There is a wide range of methods, falling in three groups, as follows, and the choice will depend on what facilities exist or can be provided.

Simulation of traffic conditions can be achieved by various kinds of functional testers and route testers. Their effectiveness depends almost wholly on the frequency at which the tests can be applied and leads to the provision of automatic routiners or their equivalent. It cannot be too strongly urged that the efficiency of any testing device should be judged separately, either for anticipation of service failure (stringent testing) or for prompt detection of service failures which have already occurred. These two requirements may conflict when both functions are performed by one tester, especially in the matter of optimum test frequency. An examination should also be made as to how many of the operational functions are tested and what portions of the exchange plant (perhaps the wiring and cabling) are not covered. Only by such critical examination can it be determined what other means of disclosing service failure may be preferable or essential.

Detection of failure of actual calls can be achieved from specific service complaints if followed by immediate tracing of the faulty route set up, but this maintenance practice becomes increasingly costly and less effective in modern complex systems. Some assistance can be given by providing visual or audible aids to call tracing and by patrolling maintenance staff or by exchange alarms, but an entirely new standard in detection of service

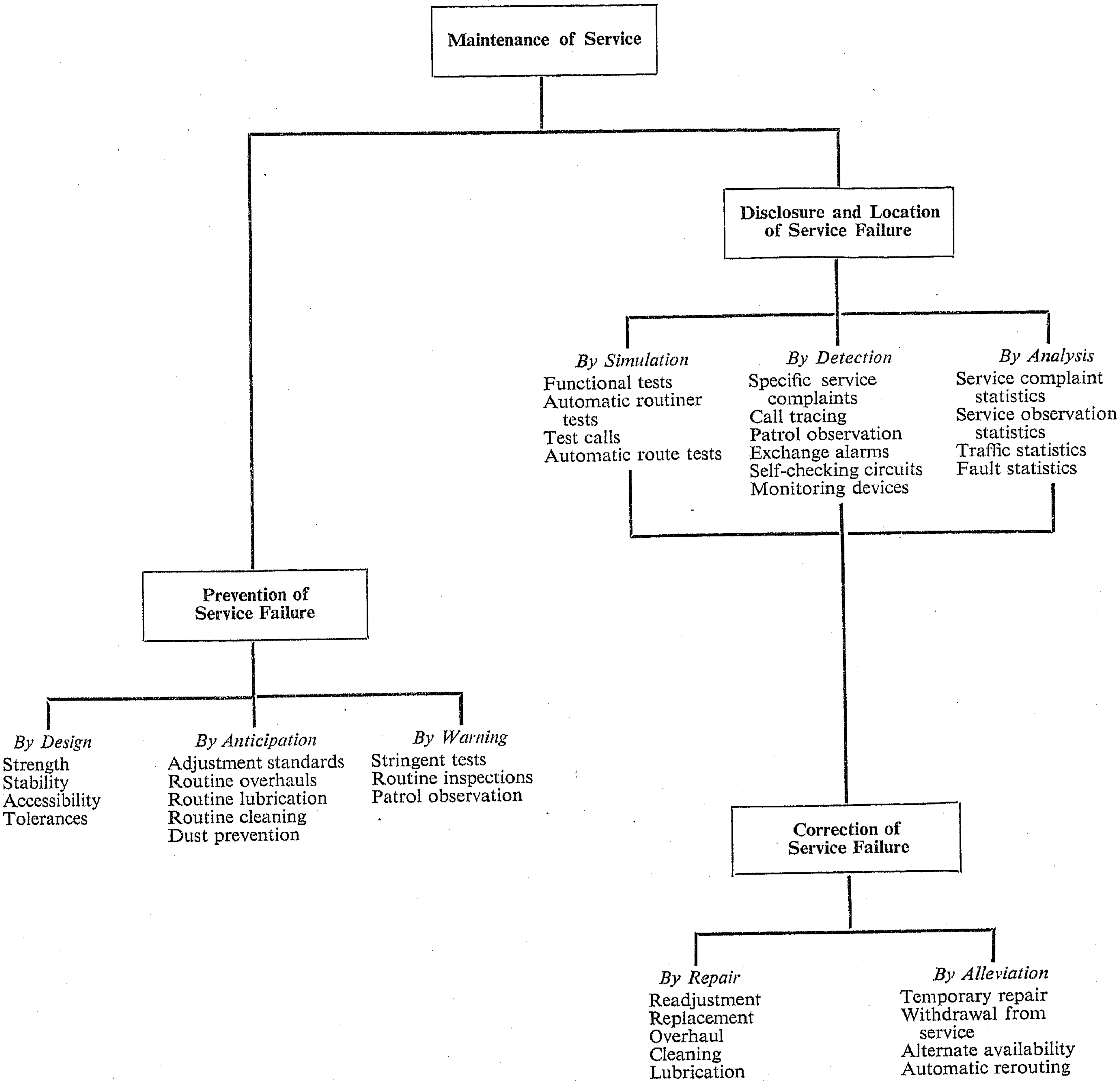


Fig. 7.—The pattern of maintenance effort.

failures can be achieved by the use of self-testing circuits and automatic monitoring devices. If the service effect of each plant fault can be reduced to, say, one-twentieth by such automatic devices, this one feature will contribute more to the maintenance of service quality than any other, but the cost of providing such facilities must be set against the maintenance advantages gained.

Analysis of statistics of faults and traffic is the third alternative method of locating trouble, and is also an overall check on the efficiency of all other maintenance work (including preventive maintenance). It may be possible to deduce from the statistics the exact location of the plant fault, but more often it provides the starting point for more precise location by simulation or observation. As an alternative to routine testing or detection by monitoring devices, the careful analysis of statistics can be developed into a very efficient procedure for local disclosure of plant faults, but it is important to use statistical procedures as a means to an end, not as an end in themselves.

(7.5) Maintenance Policies

The pattern of maintenance effort affects (or is affected by) the general maintenance policy. Both may have to be varied for common-control equipment as compared with the main selecting switches, and again may be different for individual subscriber's line circuits as compared with the common switching equipment.

A routine maintenance policy is a prominent feature of the accent on preventive maintenance, and usually includes methods

of disclosure and location of service failures by routine functional testing. Routine work has the advantage that it is more easily controlled by the higher levels of supervision, even if certain departures from a standard schedule are permitted, but it tends to focus too much attention on arrears of routines to the detriment of a more realistic attitude to quality of service. In any case, standardized routine schedules for graded switches are bound to incur unnecessary work on some selectors and to be inadequate for others.

A qualitative maintenance policy, as now adopted in the United States, is the opposite of routine maintenance in principle. The word "qualitative" implies that, whatever may be the pattern of maintenance effort, the work performed should be varied locally according to the current state of individual items of plant, the quality of the service and the skill and temperament of the technical staff. For these reasons the Americans deny the need for standard schedules of routine maintenance and, in so doing, they put preventive maintenance in its proper perspective in relation to local conditions. The qualitative principle not only stresses the need for economy in effort, but makes such economy possible.

The conclusion is therefore reached that the old-established policy of routine preventive maintenance is seriously challenged by the new approach to service quality at reasonable cost. In existing telephone exchange systems a qualitative approach enables a great deal of unnecessary work to be eliminated, with beneficial results. In new systems it is possible to incorporate self-testing and self-correcting facilities which might still further reduce the manpower required for maintenance and may at the same time make possible a higher standard of service.

In all cases the important factor is an appreciation of exactly what each maintenance function is intended to achieve. This will ensure that the efforts in any one direction are not unnecessarily stressed or duplicated, and that no important feature in the pattern of maintenance is overlooked.

(8) ACKNOWLEDGMENTS

The author wishes to thank the Engineer-in-Chief of the British Post Office for permission to publish the paper; also the American Telephone and Telegraph Company, the Automatic Telephone and Electric Co., Ltd., the British Productivity Council, and Sir Isaac Pitman and Sons, Ltd., all of whom have given permission to use data and illustrations for which they are variously responsible.

(9) REFERENCES

(1) PALMER, R. W., *et al.*: "Maintenance of Automatic Telephone Exchanges," British Productivity Council, London, 1954.

(2) STEVENSON, J. B., *et al.*: "Plant Maintenance," British Productivity Council, London, 1952.

(3) AHLSTEDT, B.: "Traffic Reliability for the L.M. Ericsson Crossbar System in Helsinki," *Ericsson Review*, 1954, p. 2.

(4) BERGLUND, C.: "Maintenance of Automatic Telephone Exchanges with 500-line Selectors," *ibid.*, 1951, p. 2.

(5) BRUBAKER, J. W.: "The Automatic Monitor," *Bell Laboratories Record*, 1950, **28**, p. 343.

(6) ELSTREE-WILSON, E. C., and LAWRENCE, C.: "The Artificial Traffic Equipment," *Post Office Electrical Engineers' Journal*, 1948, **40**, p. 159.

(7) HAGUE, A. E.: "Trouble Indicator for the Sender Link Frame," *Bell Laboratories Record*, 1946, **24**, p. 12.

(8) HANSSON, K. G.: "Maintenance of Crossbar Switch Exchange," *Ericsson Review*, 1953, p. 34.

(9) KING, W.: "Maintenance Problems in Automatic Networks," *Telecommunications Journal of Australia*, 1952, **9**, p. 28.

(10) LINDSTRÖM, E.: "New Developments on the L.M. Ericsson Telephone System with 500-line Selectors," *Ericsson Review*, 1949, p. 106.

(11) LÖNNQVIST, I.: "Jämförelse Mellan Telefontaxorna i Olika Länder," *Tele* (Swedish Edition), 1952, **1**, p. 1.

(12) MEHRING, A. C.: "Trouble Recording for the No. 5 Crossbar System," *Bell Laboratories Record*, 1950, **28**, p. 214.

(13) MODÉE, G., and EK, F.: "The Crossbar System A204 of the Swedish T. and T. Administration," *Tele* (English Edition), 1950, **2**, p. 4.

(14) MÜLLER, H.: "Betriebsüberwachung und Pflege von Fernsprechanlagen," *Fernmelde Technische Zeitschrift*, 1951, **4**, p. 294.

(15) PALMER, R. W.: "Service to Subscribers in the United States," *Post Office Telecommunications Journal*, 1954, **6**, p. 98.

(16) PREIST, T. P.: "The Case for a New Approach to Maintenance," *A.T.E. Journal*, 1954, **10**, p. 72.

(17) STONE, M. C., and GREEN, L.: "New-type Routiners for Automatic Telephone Exchange Equipment," *Post Office Electrical Engineers' Journal*, 1952, **44**, p. 147; and **45**, pp. 19 and 61.

(18) TAPLIN, C. V.: "Maintenance Center for the Crossbar System," *Bell Laboratories Record*, 1944, **22**, p. 566.

(19) URBEN, T. F. A.: "The Fault Recorder or Docket Printing Machine," *Post Office Electrical Engineers' Journal*, 1952, **45**, p. 115.

(20) WILLIFORD, O. H.: "Maintenance Facilities for the No. 5 Crossbar System," *Bell Laboratories Record*, 1950, **28**, p. 313.

(10) APPENDIX: MAINTENANCE TERMINOLOGY

Most of the terms in common use for telephone-exchange maintenance are sufficiently well understood for all practical purposes but lack precise definition for an analysis of maintenance principles. For example, all routine maintenance might be considered as preventive, yet routine lubrication is a very different function from the anticipation of the customers' fault reports by routine tests. Again, a departure from standard adjustment may be a fault that has to be corrected, yet no fault has occurred in performance. If one defective item of plant causes 100 calls to fail and leads to complaints from two separate subscribers, are there 100 faults, 2 faults, or only one?

For the purposes of the paper (and with no other authority) this kind of confusion has been avoided by using the following terms with the meanings stated. The maintenance terminology in common use for industrial plant[2] has been borne in mind, but the definitions given here are in words that apply primarily to automatic telephone exchanges.

(10.1) Definitions of Faults and Failures

Fault.—An imperfection of plant or of service, requiring attention as part of the maintenance function.

Plant Fault.—The physical deterioration or failure of a single item of plant, sufficient to require remedial action irrespective of its effect, if any, on service operation.

Service Failure.—An operational defect or failure of the service, irrespective of the plant fault that caused it.

Service Complaint.—A subscriber's report of specific service failure or other serious trouble, whether or not a plant fault is subsequently disclosed.

Breakdown.—A plant fault, or group of plant faults, causing a whole exchange or route to be out of action.

(10.2) Definitions of Maintenance Objectives

Preventive Maintenance.—The prevention of faults in preference to reliance on the disclosure and correction of faults after they have occurred. By the above definitions of different kinds of "fault," preventive maintenance can have more restricted meanings, hence—

Prevention of plant faults: purely precautionary.
Prevention of service failure: in spite of plant faults.
Prevention of service complaints: in spite of service failure.

Corrective Maintenance.—Concentration on the correction of service failures, with a minimum of preventive maintenance, and including diversion of traffic from the faulty plant.

(10.3) Definitions of Maintenance Organization

Routine Maintenance.—An organization of work for the prevention and/or disclosure of faults by means of standard schedules of regular routines to suit average conditions expected from previous experience.

Qualitative Maintenance.—A variable combination of planned work for the prevention and disclosure of faults, regulated to the local needs of plant and of service quality, and with a minimum of routine maintenance.

(10.4) Definitions of Maintenance Work

Functional Testing.—The simulation of service conditions to determine whether isolated items of plant are functioning correctly.

Stringent Testing.—The application of tests more stringent than service conditions, to check the existence of margins of safety for correct operation in service.

Route Testing.—The application of tests (functional rather than stringent) by means of artificial traffic over particular routes or in a telephone system as a whole.

Overhaul.—The physical examination and checking of isolated items of plant, mainly for the prevention of plant faults, but also for the correction of plant faults known or suspected. Overhauls are not necessarily performed to a routine schedule.

DISCUSSION BEFORE THE INSTITUTION, 10TH FEBRUARY, 1955

Mr. D. A. Barron: First, it must be stressed that, whatever principle may be favoured, it must inevitably be based on a balance and a compromise between a number of factors. Not only do some of these factors operate very often in different directions, but some are exceedingly difficult to assess and measure. The author mentions the difficulties of having a common procedure for basically different switching systems, and I shall limit my remarks primarily to Strowger equipment.

Maintenance policy is clearly related to the general policy of the telephone-operating organization, as to how good a service it wants to give or can afford to give, in relation to what it pays for its equipment, what sort of maintenance effort can be made available, what that effort costs, and what rates are charged for the service. In the framing of that policy, the need for balance and compromise immediately arises.

The next obvious compromise is in regard to the "goodness" of the equipment. If a mass-produced article is to be provided at a reasonable price, there must be tolerances and so on, and a balance has to be struck between the first cost of the equipment and what we want to get from it and what we can afford to spend in maintaining it. The author mentions as a target "reasonable service at a reasonable cost," and that is the sort of outlook which is often adopted. Some people modify the target somewhat and state, "an acceptable service at minimum cost," but I have never heard anyone maintain that it would be desirable to give the best possible service regardless of its cost, nor do I think that the subscribers would appreciate that very much.

I will give one of many possible examples to illustrate the interrelationship of the various factors. Without trying to be too specific, we may expect that a subscriber would get about 80% of his calls successfully completed at the first attempt. Of the remaining 20%, 8% or 9% at least would not be completed because the called subscriber was busy, and another 8% or 9% would not be completed because the called subscriber did not reply. This leaves only a very small percentage, which results in the main from faults and failures owing to the fact that there is insufficient equipment to carry the call at that particular time. No operating company or administration can afford to provide sufficient equipment to carry the peak traffic without any loss, since that would be quite uneconomic and a large part of the plant would not be used for the greater part of the time. It is, therefore, standard practice throughout the world to arrange that in the busiest period a very small percentage of calls will fail to mature, owing to the fact that there is not quite enough plant to carry them.

In large networks, that percentage may be of the same order as the percentage of calls which fail owing to faults. Therefore, if we want to improve the service to subscribers, it may be that by more frequent checking of plant quantities, or by having some slightly different standard of loss at each switching stage, we could, more cheaply than by maintenance improvements, give the subscriber a 0·3% or 1% improvement. He might not notice that, but if we assume that he does, we have to find the optimum method of procuring improvement, because no matter what system of maintenance is used—and the author has mentioned several—the more nearly we approach to perfection the more, disproportionately, the maintenance costs will rise.

I should like to give a few facts about some of the changes in maintenance philosophy, since the end of the war, in this country and in America. In the early days, in both countries, maintenance, whether overhaul or testing, was carried out at prescribed periods. However, in 1938 the Post Office conducted what was called the "Bristol experiment," which was one of the largest experiments they had carried out on maintenance and fault liability. As a result of the information obtained from that experiment, it was categorically stated in the national instructions that routine periodicities should be adjusted to the minimum necessary to meet proved necessity. After the war, further efforts were made towards this objective, and various guide figures were introduced to assist people in the field in determining whether they could safely reduce the periodicities of their routine overhauls and testing, and full local discretion is given. There still existed, however, the various "nominal" periodicities, and it became clear that those were some way from the average requirement, and needed revision.

In 1949, therefore, we started some controlled field experiments to discover the facts. There was a very similar change in maintenance philosophy in America in the post-war period, and I believe that in 1948 the Bell System began experiments to examine the possibilities of getting away from routine overhauls carried out at regular intervals. The similarity in thinking and timing between the two countries is remarkable, and I believe that the conclusions reached as a result of these parallel but independent investigations are essentially similar, and that the differences are more in nomenclature than in principle.

As a result of the field experiments we have found that the

average frequencies of overhauls can be considerably reduced. With some types of equipment the average intervals are now three times longer than they used to be, and the experiments are still continuing. We have, however, made changes only when we have had ample evidence from field trials to show that they were sound and fully justified in the conditions existing in this country, and having due regard to our responsibilities to the public for the maintenance of satisfactory service. In view of all the complex factors involved it is obviously desirable to move slowly and steadily, so that it has been more a re-orientation of policy than any sudden change.

The Americans have introduced what the author has described as qualitative maintenance, and I feel that it is very similar to what the Post Office have been working towards in the last five years, because the objective is the same, i.e. to do work when it is necessary, and not to a timetable. To that extent it might be said that the line of thought in the two countries is identical. There are some apparent differences. One is that, if I have understood the author correctly, the Americans tend to base the need for maintenance attention on subscribers' complaints and analyses of statistics, whereas we tend to base it on necessity based on experience, and on the results obtained from tests by artificial traffic equipments or other means, as well as on statistics. If this difference really exists, I incline to the view that our way has some advantages, in that, while it is directed towards the same objective, it leaves with the operating authority the responsibility for initiating maintenance attention in advance of appreciable service deterioration, rather than permitting any tendency for the need for maintenance attention to be, at least in part, brought to notice by a deterioration in service which has resulted in an increase in complaints.

Routine testing is not quite the same, but there are some similarities in principle, and we are moving towards more through-testing and less routine limit-testing. However, although we have reduced the amount and, to a degree, the nature, of automatic routining, we do not propose, especially for common equipment, to eliminate it.

Mr. W. H. Grinsted: I have never been satisfied in having to offer customers elaborate routine-testing gear thought to be necessary to keep the exchange equipment working, and I have been still less satisfied in having to manufacture that equipment, or the numerous components wanted for it, in "one-off" quantities by mass-production methods and sell them at mass-production prices. The manufacturer would very much prefer to be able to recommend what the author describes as the "qualitative maintenance policy" for his equipments.

In order to be able to do that there are definitely two requirements which have to be met. The first is that the equipment must be really reliable, and the second is that the ambient conditions in the exchanges must be satisfactory. With regard to reliability, the author refers in Section 2.2 to the stability of the apparatus and of its adjustment, and in Section 2.3 to what for the sake of brevity may be called the "circuit margins." Fault statistics, as at present compiled, afford very little opportunity to study the shortcomings of the equipment in those two respects. An analysis showing the number and nature of faults arising from shortcomings in those directions would be very useful to the manufacturers. I should like to know whether the author thinks there is any possibility of issuing statistics compiled in such a way as to show the failures which can be referred to lack of stability or unreliability in the apparatus design, and to too small margins in the circuits.

In the early days of automatic equipment, I was able to collect a number of figures which pointed to the principal causes of unreliability as being relay contacts, mechanical spring sets, wiper cords and wipers. So far as relay contacts are concerned a very great improvement was effected by the introduction of the twin relay contact, but very little has been done in the way of improvement in mechanical spring sets or in wipers and wiper cords, at any rate in the Post Office standard system.

Circuit margins are a much more complicated subject, but I feel that, since customers have undertaken their own circuit design, a great deal of the "no-man's-land" which ought to lie between the demands of the circuit and the limiting performance of the apparatus has been annexed by the circuit designer.

In connection with the conditions in the exchanges, it has seemed that, since the introduction of the twin relay contact, there has been some deterioration in the precautions which were taken in the early days against contamination of the equipment by dust and dirt. The twin relay contact effected such an improvement that the troubles due to dust experienced with the earlier equipments more or less disappeared, and since then we have perhaps been a little too lax in the precautions taken to protect the equipment. I agree with the author's remarks on dust and ventilation, and his comments should be taken very seriously.

There is one sentence which particularly appealed to me in Section 6.5, where the author states that with qualitative maintenance the local staff are required to analyse the statistics and other indications to justify almost every maintenance function that they undertake. In other words, they are required to ask three questions asked by any good engineer before he incurs expenditure: Why do this at all? Why do it now? Why do it in this way? Contrast that with the performance of routines just because they are routines and we have complete justification for the policy of qualitative maintenance.

The author emphasizes that it is the service which has to be maintained, and that maintenance of the plant is only a means towards that end. As a rough and ready method of making economic calculations, we used to allow a percentage on the capital cost of the plant to represent the cost of maintenance, and even with "work unit" calculations, plant seems to be the basis of the assessment. Is there not a case for trying to make an assessment for each item of the equipment depending on its contribution to the service, perhaps based on the number of calls which it handles or some other more suitable figure, so that targets for maintenance expenditure can be allocated on a basis of service rather than of plant?

Mr. T. H. Flowers: I would like to refer first to the fact that there are no numerical data in the paper by which the effects of various maintenance procedures may be judged. The lack of accurate measurement of the effects of maintenance may account for the contradictory experiences reported, e.g. at the beginning of Section 3.2 and later in the same Section, where it is recorded that recent tests have suggested that some kinds of preventative maintenance may do more harm than good. The difficulties of making measurements in this field are great, but not more so than in other sciences, biology for example. The cost of large-scale experiments is small compared with the annual expenditure on maintenance and could be recouped by quite small improvements in practices.

In the first paragraph of Section 2.1 it is stated that fault location is easier in systems without, than in systems with, common apparatus, but I doubt whether the substantial portion of the world which uses common apparatus would agree. There are faults which consistently recur every time that certain apparatus is involved and those which do not. Those in the first class, e.g. wiring disconnections, are easily discovered in any system. The second class includes intermittent faults and faults which depend on particular conditions such as the selection of particular outlets. The functional testers, routiners and route testers are successful with the second class of fault only if

by chance they happen to encounter the intermittent fault or faulty outlet. The self-testing and automatic monitoring devices mentioned in the paper will, in time, find such faults as and when they cause trouble on particular calls. Devices of this kind can, however, be economically provided only in common apparatus systems. The overall difficulties of fault finding thus appear to be dependent on the proportions of consistent and inconsistent faults. I have no data on this matter, but certainly normal wear and deterioration are very likely to produce intermittent faults before they have progressed far enough to produce consistent faults. On the balance of experience, I think that the problem of fault finding can always be satisfactorily solved technically whatever system is adopted. The cost involved varies with different systems, but it is only one item among several in determining the total cost of service, and it is total cost which is important.

In a future system it is possible that there will be some electronic devices, and for these many of the entries in Fig. 7 will no longer apply. Electronic devices cannot have routine cleaning, dust prevention or lubrication, and it is no use having patrol observation, because it is not usually possible to see what is happening. Hence the "pattern of maintenance effort" will be simpler, and I think that the actual maintenance of an electronic telephone exchange would also be simpler than that of the present types.

Mr. T. P. Preist: The author's definition of a fault is too wide and leaves too much to the personal assessment of the maintenance officer. I prefer to define a fault as "any defect which prevents a circuit or a piece of apparatus from fulfilling its function." On this definition, a change of condition which does not prevent the operation of the equipment, such as the departure from some specified standard adjustment, is not a fault.

I agree that acceptance tolerances should be as tight as possible, but they should be checked by functional testing only, and not by a physical check of adjustments, since this depends too much on the tester's interpretation of right and wrong. To illustrate this a spring was tensioned to 47 grammes (accurately measured by weights), and six people were asked to measure it with a standard tension gauge. Their results varied from 43·5 to 51 grammes. This variation of assessment causes much damage because each person makes changes to suit his own interpretation. Stringent functional testing is preferable for acceptance testing, since no physical changes are made unless they are shown to be necessary.

The evidence in the paper for and against regular overhauls is contradictory. Our experience, particularly oversea, shows that overhauls are only worth while when the staff is well selected, expertly trained, interested, and has time to do the work properly. If this is not the case, more harm than good is done, and frequent functional testing to limits very slightly in excess of exchange requirements is preferable to overhauls.

I consider that routiners have been wrongly used in the past. They have been designed to test to the optimum limits, and whether used for acceptance tests, checking after a major repair or regular overhaul, or for day-to-day testing, no change has been made to those limits. I suggest routiners be arranged, first, to make stringent tests for acceptance, or tests after a major repair, secondly, to apply reduced limit tests for regular interval testing, thus making allowance for the natural variation of adjustment that the switch can tolerate, and thirdly, to apply no-limit tests for day-to-day testing in order to ensure that the equipment will perform its normal functions.

Artificial traffic equipment is very useful, particularly for service observations, and I find that the results obtained are often worse than those shown by manual service observations.

With reference to the question of whether to overhaul or not, the organization with which I am associated made some laboratory experiments which showed that the policy of not overhauling had much in its favour. The results, which have been published in some detail, have been mildly criticized on the ground that laboratory experiments cannot wholly simulate exchange conditions. The results were, however, highly significant, and we would welcome a large-scale field experiment by the Post Office on similar lines.

Col. C. E. Calveley: I regard this problem as primarily a management and not a technical one. The higher level of management must decide what proportion of the total effort on maintenance is to go on to the floor of the exchange, how much is to go into examination of results of routine tests and the analysis of faults, so that guidance may be given to the engineer in charge of the exchange and plant weaknesses may be brought to light. The engineer-in-charge also has the problem, with a given staff, of deciding how to direct the efforts of his staff to those items of the plant which past and present experience indicate require attention.

On the first point, I wonder whether the author has any information comparing the number of men on the staff (in the military sense) with those in the line organization in British and American practice. It was my impression that the American organization employed far more people on staff duties than we do in this country.

With regard to the second matter, there is a larger number of factors operating. First, there is the variation in the total operations of the switch in a year depending on its position in the grading as shown by the author in Fig. 2. Secondly, there is the stabilizing period of new equipment; in an analysis of a particular exchange the number of faults per group selector in the first three years was 2·8, 3·4 and 0·7. Thirdly, there is the variation between exchanges, since the fault rate in an equivalent group selector in another exchange was 0·48, 0·11 and 0·18. Fourthly, there is the ability of a switch to give satisfactory service, although some of its adjustments are outside the tolerances previously found to be desirable; for example, it was found that there was no mechanical clearance between the rotary detent and the long face of the ratchet after about 50 000 operations, but the first fault in service through this cause did not occur until after a further 127 000 operations had taken place.

The engineer in charge of the exchange has to be provided with data which will enable him to take the above and other factors into account in directing the efforts of his staff to ensure that the best service is given to the public.

With our limited experience of electronic equipment in service at Richmond, we have found that the number of faults bear little relation to the amount of use. Four faults have occurred on an electronic director which has handled 628 000 calls in 18 months, whereas eight have occurred in the same period on a late-choice electronic director which has only handled 7 400 calls.

The author has not mentioned the contribution that the interest and enthusiasm of the local staff can make. It is a possible explanation of the improvement in service which has been obtained in different exchanges by contradictory methods. The solution to the problem of arousing, developing and maintaining enthusiasm and interest would probably change the system of maintenance effort and produce better results, but this is a management and not a technical problem.

Mr. N. C. Smart: In Section 4.1 the author states that gritty dust increases the rate of wear, especially if combined with oil or grease. I thought that surfaces which rubbed together were lubricated in order to stop wear. Whether dust is present or not, we still lubricate to stop wear. How is it that with a wiper and bank the opposite effect is obtained? When I pose this question it is stated that a type of grinding paste is formed, but if we want to lap two surfaces together we put abrasive powder between

them, and to stop it scoring, i.e. wearing too quickly, we put oil into the powder and make it into a paste.

The general background of the paper and discussion makes the point that everything requires a little too much maintenance. Perhaps, however, there is an excuse for looking at the other side of the picture. Is any of the apparatus too good? The subscribers' uniselector was designed to have a life of 25 million steps, which represented one million calls. Now we have a double homing bank, so that we have only 12 steps per call, which means a life of two million calls. I do not know what that represents in years, but it must be quite fantastic. Is it possible to design a cheaper switch which would give adequate service?

I was impressed by the fact that the only routiner I saw in American exchanges was a subscribers' line routiner, and it was earning very high praise from the maintenance men and the management. It is surely strange that the only thing that we do not routine test in this country should be the only thing that they do routine test there. Has the author had any experience of that routiner?

Mr. S. Rudeforth: Most of us, including the Americans, are agreed that the main problem is to obtain service out of plant at minimum cost. Controversy arises in translating the various ideas into practice to get that minimum cost.

Plant, no matter how perfect it may be thought to be, is liable to random faults. In telephone equipment faults arise from dirty contacts, broken wires or broken parts. On the incidence and duration of such faults will depend the quality of service given.

If we concede that apparatus is subject to random faults, we must locate them by routining at such a rate that the service does not fall below a stipulated value. We should fix the standard of service and then try to achieve it at minimum cost. If we functionally routine at a rate governed by the number of routine operations performed per faulty item found, we have a feedback system by which we can adjust the rate of routining to keep the number of routine operations per faulty item at some value related to the standard of service desired. Thus routining can be adjusted to the minimum requisite to meet a given standard of service and dependent on the particular conditions under which the plant has to work.

The author's definition of qualitative maintenance agrees very closely with what is standard procedure in the Post Office, but is not quite the same thing as I understand the American so-called qualitative system of maintenance to be. This seems to involve stopping routines altogether, waiting for the subscribers to complain, analysing the causes of their complaints and then dealing with the equipment. Long-term efficient local control demands a feedback system of the kind described.

With regard to overhauls, switches, like individuals, are all different. Even when switches are subject to exactly the same number of operations, there will be a distribution from the good to the bad. Differences between switches are accentuated because, in fact, they do not carry the same traffic. If you treat all the switches alike and overhaul them at regular intervals, you will overhaul some unnecessarily and may fail to deal with those which require overhaul earlier. We have found from our experiments started four years ago that within limits the longer we can leave switches between overhauls the better performance we get from them. On the other hand, advice to leave switches indefinitely has to be qualified because sooner or later they will need attention, and to leave the whole exchange until it wants a mass attack is not good maintenance practice; this involves a regular minimum of controlled work to keep the service at a desired value. The logical thing to do, therefore, is to treat switches as individual personalities and to overhaul them only when they require it; in general, this is when they fail because of faults due to wear and tear.

Mr. W. H. Scarborough: The author has mentioned qualitative maintenance, and I can quote two or three cases. About 1934 we were having a good deal of trouble at the Bishopsgate exchange, and the difficulty was traced to L.1. contacts of the first code selectors. At the time we were developing automatic working at the City, Central, Waterloo and other exchanges, and there was a shortage of staff. We were advised not to stop routining. We had to detail men to overhaul the first code selectors. We stopped all routining, except bank cleaning and wiper inspections, and used some operators during the busy period to pass 200 calls per day, on selected days, through the equipment, to trace faults. We had no trouble, and the maintenance for 12 months was as good as that of any comparable exchange, if not better.

In a recent case a new exchange in the first 6 months showed about 3% plant faults. I said that this was ridiculous and we must bring the faults down to less than 1%. We decided to use operators to pass calls through the equipment during the busy period, and we found the faults. In another case when there was trouble in the trunk exchange, we found that the faults were not in the 2-V.F. equipment of that exchange but in the 12½-mile London network. We pumped calls into the network and found the faults. This is by far the best way to find faults, and therefore I am wholly in favour of qualitative maintenance.

Mr. G. A. Probert: We are all agreed on the need for clean air in our telephone-exchange apparatus rooms. Therefore, why do we continue to put windows which can be opened in these rooms? Windows which open cost more, and even when closed permit the entrance of dust-carrying draughts. The Post Office has a large building programme in hand, and now is the time to attend to this matter.

Accessibility of equipment, as the author states, is an important maintenance consideration. The present method of rack-mounting the Post Office No. 3 uniselector, 25 to a row, with the lowest row fixed 11¼ in from the floor is an example of bad accessibility and should be corrected. Uniselectors have to be adjusted *in situ*, and this can only be done in the case of the bottom row by lying down, as the eye level must necessarily be within a foot of the floor. With regard to accessibility, I think we should where possible arrange the equipment so that that which has to receive the greatest maintenance attention is easily accessible. For example, P.B.X. subscribers' uniselectors get up to ten times the wear of those of ordinary subscribers, and I consider that we should as a standard practice arrange that P.B.X. uniselectors occupy the middle shelves of the racks.

Reference has been made to the long life of uniselectors, but they do need maintenance attention; banks have to be cleaned and adjustments made *in situ*, which is a very tedious job. Experience indicates that towards the end of their life, i.e. after 15–20 years, maintenance becomes expensive. The Post Office originally had uniselectors, then they changed to line finders and now back again to uniselectors. Have we not made a wrong move in going back to uniselectors? I think that a line-finder system without secondary working would be better in the long run.

Sir Gordon Radley: Two maintenance policies run through the paper, although perhaps the second comes out strongly only at the end. One is the traditional policy which prescribes schedules for tests, inspections and overhauls at regular intervals. Perhaps when we started we erred on the side of caution and made those intervals too short, but the Post Office has recently been adjusting them in the light of experience; Mr. Rudeforth's feedback analogy was apt. We have also given not only freedom but, indeed, responsibility to the man on the spot to vary routine frequencies according to local conditions.

In America, the policy of qualitative maintenance is being tried. This means leaving well alone and giving attention to the

equipment only when the service indicates that it is necessary. It demands statistical observation of the quality of service.

These two policies have been hotly debated, but Mr. Barron has queried whether they are not really the same. However, I am quite sure that, provided that routine frequencies are flexible and adjusted in the light of experience, both policies may be capable of giving equally good results.

Mr. A. J. Litton (*communicated*): At the end of Section 2.1 the author implies that the fault liability on switches is in direct ratio to the traffic carried. It is thought that this is only very broadly true and that certain faults, e.g. those due to dirty contacts and banks, are more apt to occur on switches which are rarely used.

The cost and value of preventive maintenance for automatic-exchange equipment has been closely scrutinized in Ireland for some time. As a result the policy has been changed and it now resembles somewhat the American qualitative maintenance policy. The overhauling of switches has been found to be expensive in man-hours and has shown negligible returns in reducing the fault liability, and, in fact, it frequently reduced the life of working parts. Therefore, apart from periodic cleaning and lubrication (carried out so far as possible with a switch *in situ*) no overhauls are now carried out.

The automatic routiner has, in the past, been the main means of finding faulty switches. However, it indicates faults which (*a*) do not affect the service or (*b*) are self-clearing in a relatively short time, and the cause of which it is practically impossible to find. Moreover, the staff time taken by routines and the investigation of the "faults" disclosed is appreciable.

Thus the trend in Ireland has been to make more use of patrol duties, "cutting in" tests and artificial traffic equipment. Tests with automatic routiners which were formerly made weekly are now only done at two-monthly intervals. The change in the distribution of faults found by the various methods is shown in Table A.

At the same time the switch fault index has fallen from 1·0 to 0·6 fault per switch per annum without any deterioration in service.

Automatic-to-automatic relay sets are still routined automatically once per week (since no other suitable tests exist). In this case it has been found that the testing can be speeded up by modifying the routiner so that by throwing a key a number of the tests which seldom show up faults can be stepped over for normal routines, the full routine only being carried out occasionally.

Table A

Faults found by	Percentage of total faults	
	Routine tested at weekly intervals	Routine tested at 2-monthly intervals
	%	%
Routiner ..	50	10
Cut-in tests ..	35	55
Artificial traffic, patrol, etc.	15	35

Statistics obtained from the equipment fault records and from the engineering repair service have shown that the service has not deteriorated; at the same time maintenance costs have decreased by approximately 20%. The fault statistics have also been used to indicate where special action or investigation is required on any section of the plant.

The artificial traffic equipment has been found to be a most useful tool in assessing the quality of service, since calls are made under controlled conditions and faults can readily be traced when they occur; it is considered to be, in general, much more satisfactory than the usual service observation. As mentioned in Section 6.3, it is sometimes difficult to get sufficient traffic on subscribers' circuits to make service observation worth while; in such cases observations can usefully be made from first group selectors.

The author's remarks on cleanliness are most interesting. It is felt that much more consideration might be given to enclosing switches and banks under suitable covers as is usual in Continental practice. The extra cost would be covered many times over in the maintenance charges.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. R. W. Palmer (*in reply*): The uniselector versus line-finder controversy mentioned by Mr. Probert is an example of choice of equipment design being dependent to a large extent on maintenance efficiency. It is true, as Mr. Smart says, that cheaper apparatus might be possible, but the ample margin of reliability of the British uniselectors may well be the reason why they are better than line-finders for large exchanges in this country. Similarly, the need for accessibility on equipment racks is dependent on the frequency of maintenance attention demanded by the apparatus design and by the maintenance policy that is practised.

I agree with Mr. Litton that fault liability is not always proportional to traffic. The point intended in Section 2.1 was that the traffic carried by any particular selector is a direct measure of the effect of any plant fault on the overall service. Hence the effort expended on looking for faults on particular selectors should be proportional to the amount of traffic that may be affected.

On the subject of dust prevention, it may be impracticable to prohibit windows that open, as Mr. Probert would like, but I fully support Mr. Litton's appeal for some simple and cheap cover for selector banks. The initial trials might have to be with an expensive type of cover, merely to determine what maintenance savings can be achieved. This could then be compared with the capital cost of moulded plastic covers, assuming that they would be manufactured in large quantities.

The subscribers' line routiner mentioned by Mr. Smart is limited to one predictable fault, namely deterioration of insulation resistance. In the American view, the subscriber will know and report when his line fails (in the operational sense) far more promptly than any practical routine test that could be made by the telephone company. This is an example of the use of a routiner exclusively for the prevention instead of for the disclosure of service failures.

Mr. Rudeforth's suggestion that qualitative maintenance involves stopping routines altogether refers only to the launching of this policy on an experimental basis. The essential feature of the American outlook is that the subsequent wide variation in frequency of the routines that may be re-introduced is of no consequence whatever so long as the quality of service and the maintenance costs are satisfactory. Several speakers also misunderstood qualitative maintenance as being dependent on an increase in the number of complaints from subscribers. Surely we must admit that, on the average, the quality of service in the United States is no worse than in the United Kingdom. We must also admit that even under a routine or preventive maintenance policy the lost calls and the subscribers' complaints

are by no means negligible and provide ample data for "qualitative maintenance" action, without waiting for more to be incurred.

Within the scope of an Institution paper it has been impossible to present the statistical data that Mr. Flowers and others would have wished. All that could be attempted has been the presentation of the range of factors involved and some principles for their classification. When statistics are required on some particular factor, whether it be maintenance cost per call, the stability of apparatus or the effect of some maintenance procedure or policy, it is at least an essential starting-point to be aware of the unwanted variables. Col. Calveley and Mr. Preist reminded the meeting that one of these variables is the human factor. In my own experience, the influence of foremanship as practised in the United States and Sweden can swamp many of the variations in technical policy and procedure.

It is for such reasons that the success of maintenance techniques can be determined only by large-scale field trials. I cannot support too strongly the views on this point expressed by Messrs. Barron, Flowers and Preist. The Bell System in America may have plunged into bolder experiments, true to their national character, whereas the British Post Office has followed the more analytical and cautious route; but, as Sir Gordon Radley so rightly implies, the conclusions should be the same. It is the practical result which counts, and particularly the result as seen by the subscriber.

621.385.4

Paper No. 1835 R
July 1955

AMPLIFICATION FACTORS AND MUTUAL CONDUCTANCE OF A BEAM POWER VALVE

By S. DEB.

(*The paper was first received* 16*th June, and in revised form* 13*th December*, 1954).

SUMMARY

Analysis has been carried out for the determination of the distribution of electric field in a beam power valve when the control-grid pitch is larger than the cathode–control-grid distance. The method followed is an extension of the one suggested by Fremlin for a triode valve. Expressions have been derived for the various amplification factors, μ, and it is shown that, in general, these factors vary along the surface of the cathode. It is further shown that the earlier expressions that have been derived give only the average values of the amplification factors.

Expression for mutual conductance, g_m, has also been considered for the beam tube. It is shown that, by taking account of the so-called *inselbildung* effect, arising out of variation of μ referred to above, close agreement between theoretically computed and experimentally obtained characteristics may be secured.

(1) INTRODUCTION

A highly efficient and widely used power valve is the so-called beam tetrode, in which the undesirable effects of secondary emission from electrodes are eliminated by using aligned grids, deflecting plates and large separation between the screen grid and the anode. The underlying principle of operation of such valves is conveniently studied by considering separately the two distinct regions of the valve, namely:

(*a*) the region between the screen grid and the anode, and
(*b*) the region between the cathode and the screen grid.

Region (*a*) is investigated by studying closely the nature of the flow of the space charge. The investigation leads to an understanding of the shape of the plate characteristic curve and the efficiency of the valve. Region (*b*) is best studied by analysing the distribution of the electric field. The analysis leads to an understanding of the mutual characteristics of the valve.

The problem connected with (*a*) has been treated in detail for a valve of plane geometry, and the results obtained serve as a useful basis for the design of beam power valves.[1,2] With regard to (*b*), direct analysis has been carried out for plane valves for the case when the grid-wire radii are small and inter-electrode distances large compared with the grid pitch.[3] The results obtained, however, are often only partially successful in elucidating the characteristics of beam power valves as actually manufactured. This is mostly due to the fact that the distance from cathode to control grid of such valves is generally smaller than the grid pitch. For valves having such a geometry the amplification factor, and hence the cut-off bias, might generally vary with position in relation to the grid wires. Under this condition, over a portion of the negative grid region of the characteristic curve, emission of current might be confined to isolated strips or islands on the cathode surface giving rise to the phenomenon known as *inselbildung*. Hence, the characteristic curves computed from the results of simple analysis referred to above deviate more and more from the experimental results as the cut-off point is approached. When the cathode–control-grid distance is not less than one-half the grid pitch an indirect method, suggested by Liebmann[4] and based on the results of his analysis of electric field in a triode, is helpful in computing the amplification constants and mutual characteristics of a beam power valve. A direct and less restricted analysis of the electric-field distribution in a beam tetrode for the case of small cathode–grid distance has not hitherto been made. This has been done in the present paper by extending the method used by Fremlin for triodes of small cathode–grid distance. The analysis is found to yield very good agreement between theoretically computed and experimentally obtained characteristics of the valve.

(2) ELECTRIC FIELD IN A BEAM POWER VALVE

We consider the case of a "plane" valve in which the cathode and the anode are plane surfaces and the grids consist of sets of equidistant straight parallel wires. A cross-section of the valve is shown in Fig. 1, where C, g, S and A indicate respectively the cathode, the control grid, the screen grid and the anode. The symbols used for the various dimensions of the electrode structure are shown in Fig. 1.

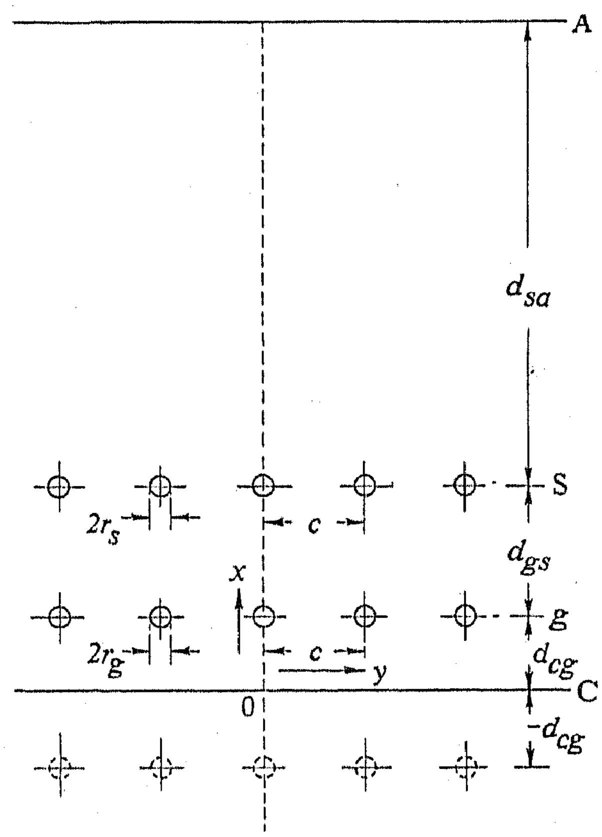


Fig. 1.—Cross-section of a plane beam power valve.

C = Cathode.
g = Control grid.
S = Screen grid.
A = Anode.

The image of the control grid, shown by the broken line below C, must be taken into account when $d_{cg} < c$.

For the purpose of analysis the y-axis will be taken as coincident with the cathode section and the x-axis as passing through one

Written communications on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Deb is at the Institute of Radiophysics and Electronics, University College of Technology, Calcutta.

of the grid wire centres. If the space-charge effect and end effect are neglected, the potential at any point in the valve is given by the sum of the potentials due to the charges on the valve elements, namely the line charges on the two grids and the charges on the anode and cathode surfaces, the boundary condition being that the potential has a constant value over the surfaces of these elements.[5] Thus, we have for the case of small screening constant of the grids

$$V = -q_g \log 2\left[\cosh\frac{2\pi}{c}(x-d_{cg}) - \cos\frac{2\pi y}{c}\right] - q_s \log 2\left[\cosh\frac{2\pi}{c}(x-d_{cs}) - \cos\frac{2\pi y}{c}\right] - Bx + K \quad . \quad . \quad . \quad . \quad (1)$$

where q_g, q_s, B and K are constants to be determined from boundary conditions. This simple expression for V, however, ceases to hold if

$$\frac{d_{cg}}{c} < 1$$

as in this case, the expression does not give constant value of the potential over the cathode. In the analysis that follows, this limitation will be overcome by introducing in the expression for V, the potential due to an image of the control grid, as has been done by Fremlin in studying triodes of small cathode–grid distance.[6] We then have

$$V = -q_g \log \frac{\cosh\frac{2\pi}{c}(x-d_{cg}) - \cos\frac{2\pi y}{c}}{\cosh\frac{2\pi}{c}(x+d_{cg}) - \cos\frac{2\pi y}{c}} - q_s \log 2\left[\cosh\frac{2\pi}{c}(x-d_{cs}) - \cos\frac{2\pi y}{c}\right] - Bx + K \quad . \quad (2)$$

If $V = 0$ at $x = 0$, i.e. if the cathode is assumed to be at zero potential, we have

$$K = q_s \log 2\cosh\frac{2\pi d_{cs}}{c} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

For the grid-wire surface we consider the point (d_{cg}, r_g) and write $V = V_g$.

Hence
$$V_g = -q_g \log \frac{1-\cos\frac{2\pi r_g}{c}}{\cosh\frac{4\pi d_{cg}}{c} - \cos\frac{2\pi r_g}{c}} - q_s \log\frac{\cosh\frac{2\pi d_{gs}}{c} - \cos\frac{2\pi r_g}{c}}{\cosh\frac{2\pi d_{cs}}{c}} - Bd_{cg} \quad . \quad (4)$$

Assuming that d_{gs} and $d_{cs} \gg c$ and noting that

$\log 2\cosh x \simeq x$ (when x is large)

and $\sin\theta \simeq \theta$ (when θ is small)

eqn. (4) becomes

$$V_g \simeq -q_g\left(2\log\frac{2\pi r_g}{c} - \frac{4\pi d_{cg}}{c}\right) - q_s\left(-\frac{2\pi d_{cg}}{c}\right) - Bd_{cg} \quad . \quad . \quad . \quad . \quad (5)$$

Proceeding in the same fashion we get

$$V_s \simeq -q_g\left(-\frac{4\pi d_{cg}}{c}\right) - q_s\left(2\log\frac{2\pi r_s}{c} - \frac{2\pi d_{cs}}{c}\right) - Bd_{cs} \quad . \quad (6)$$

$$V_a \simeq -q_g\left(-\frac{4\pi d_{cg}}{c}\right) - q_s\left[\frac{2\pi}{c}(d_{sa} - d_{cs})\right] - Bd_{ca} \quad . \quad (7)$$

where V_s = screen-grid potential, and
V_a = anode potential.

Eqns. (5), (6) and (7) can be written as

$$\left.\begin{aligned} a_1B + b_1q_g + c_1q_s + V_g &= 0 \\ a_2B + b_2q_g + c_2q_s + V_s &= 0 \\ a_3B + b_3q_g + c_3q_g + V_a &= 0 \end{aligned}\right\} \quad . \quad . \quad . \quad (8)$$

Solving by the method of determinants,

$$\left.\begin{aligned} B &= -\frac{\Delta_B}{\Delta} \\ q_g &= -\frac{\Delta_g}{\Delta} \\ q_s &= -\frac{\Delta_s}{\Delta} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where

$$\left.\begin{aligned}
\Delta_B \simeq{} & V_g\frac{8\pi d_{cs}}{c}\left(\log\frac{2\pi r_s}{c} - \frac{\pi}{c}d_{sa}\right) \\
& - V_s\left[\left(2\log\frac{2\pi r_g}{c} - \frac{4\pi d_{cg}}{c}\right)\frac{\pi}{c}(d_{sa}-d_{cs}) - \frac{4\pi^2}{c^2}d_{cg}^2\right] \\
& + V_a 4\left[\log\frac{2\pi r_g}{c}\log\frac{2\pi r_s}{c} + \frac{\pi}{c}\left(\frac{\pi}{c}d_{cg}d_{gs} - 2d_{cg}\log\frac{2\pi r_s}{c} - d_{cs}\log\frac{2\pi r_g}{c}\right)\right] \\
\Delta_g \simeq{} & -V_g 2\left(\frac{2\pi}{c}d_{sa}d_{cs} - d_{ca}\log\frac{2\pi r_s}{c}\right) \\
& + V_s\left(\frac{4\pi}{c}d_{sa}d_{cg}\right) - V_a\left(2d_{cg}\log\frac{2\pi r_s}{c}\right) \\
\Delta_s \simeq{} & V_g\left(\frac{4\pi}{c}d_{sa}d_{cg}\right) - V_s 2\left(\frac{2\pi}{c}d_{ga}d_{cg} - d_{ca}\log\frac{2\pi r_g}{c}\right) \\
& + V_a\left(\frac{4\pi}{c}d_{cg}d_{gs} - 2d_{cs}\log\frac{2\pi r_g}{c}\right) \\
\Delta \simeq{} & \frac{8\pi d_{cg}^2}{c}\left(\log\frac{2\pi r_s}{c} - \frac{\pi}{c}d_{sa}\right) \\
& - 2d_{cs}\left[\left(\log\frac{2\pi r_g}{c} - \frac{2\pi d_{cg}}{c}\right)\frac{2\pi}{c}(d_{sa}-d_{cs}) - \frac{4\pi^2}{c^2}d_{cg}^2\right] \\
& + 4d_{ca}\left[\log\frac{2\pi r_g}{c} + \frac{\pi}{c}\left(\frac{\pi}{c}d_{cg}d_{gs} - 2d_{cg}\log\frac{2\pi r_s}{c} - d_{cs}\log\frac{2\pi r_g}{c}\right)\right]
\end{aligned}\right\} \quad (10)$$

Eqns. (7), (9) and (10) determine the electric-field distribution inside the beam power valve.

(3) AMPLIFICATION FACTOR AND CAPACITANCE COEFFICIENT

From eqn. (2) we obtain the following expression for the off-cathode potential gradient:

$$\left(\frac{\delta V}{\delta x}\right)_{x=0} = q_g\frac{4\pi}{c}f(y, d_{cg}) + q_s\frac{2\pi}{c}f(y, d_{cs}) - B \quad . \quad (11)$$

where

$$f(y, d_{cg}) = \frac{\sinh \frac{2\pi d_{cg}}{c}}{\cosh \frac{2\pi d_{cg}}{c} - \cos \frac{2\pi y}{c}} \quad . \quad . \quad . \quad (12)$$

$$f(y, d_{cs}) = \frac{\sinh \frac{2\pi d_{cs}}{c}}{\cosh \frac{2\pi d_{cs}}{c} - \cos \frac{2\pi y}{c}} \quad . \quad . \quad . \quad (13)$$

Since $d_{cs} \gg c$

$$f(y, d_{cs}) \simeq \tanh \frac{2\pi d_{cs}}{c} \simeq 1$$

Hence, $$\left(\frac{\delta V}{\delta x}\right)_{x=0} = q_g \frac{4\pi}{c} f(y, d_{cg}) + q_s \frac{2\pi}{c} - B \quad . \quad . \quad (14)$$

An examination of the relative influences of the various electrodes upon the off-cathode gradient leads readily to the various amplification constants of the valve. The two amplification factors that are of importance in a beam tetrode are (*a*) the amplification factor of the control grid with respect to the anode (μ_a), and (*b*) the amplification factor of the control grid with respect to the screen grid (μ_s).

If L/Δ, M/Δ and N/Δ denote respectively the coefficients of V_g, V_a and V_s in the expression for off-cathode potential gradient, eqn. (11),

$$\mu_a = L/N$$

and $$\mu_s = L/M$$

From eqns. (10) and (12) we obtain

$$\left.\begin{aligned} L &= \frac{8\pi}{c}\left[d_{cg}\left(\frac{2\pi}{c}d_{sa} - \log\frac{2\pi r_s}{c}\right) - f(y, d_{cg})\left(\frac{2\pi}{c}d_{sa}d_{cs} - d_{ca}\log\frac{2\pi r_s}{c}\right)\right] \\ M &= \frac{16\pi^2}{c^2}d_{sa}d_{cg}f(y, d_{cg}) - \frac{8\pi}{c}d_{sa}\left(\frac{2\pi}{c}d_{cg} - \log\frac{2\pi r_g}{c}\right) \\ N &= 4\log\frac{2\pi r_s}{c}\left(\frac{2\pi d_{cg}}{c} - \log\frac{2\pi r_g}{c}\right) - \frac{4\pi}{c}f(y, d_{cg})\log\frac{2\pi r_s}{c} \end{aligned}\right\} \quad (15)$$

We thus get

$$\mu_a = \frac{d_{cg}\left(\frac{2\pi}{c}d_{sa} - \log\frac{2\pi r_s}{c}\right) - \left(\frac{2\pi}{c}d_{sa}d_{cs} - d_{ca}\log\frac{2\pi r_s}{c}\right)f(y,d_{cg})}{\frac{c}{2\pi}\log\frac{2\pi r_s}{c}\left(\frac{2\pi d_{cg}}{c} - \log\frac{2\pi r_g}{c}\right) - d_{cg}\log\frac{2\pi r_s}{c}f(y, d_{cg})} \quad . \quad . \quad . \quad . \quad (16)$$

$$\mu_s = \frac{d_{cg}\left(\frac{2\pi}{c}d_{sa} - \log\frac{2\pi r_s}{c}\right) - \left(\frac{2\pi}{c}d_{sa}d_{cs} - d_{ca}\log\frac{2\pi r_s}{c}\right)f(y,d_{cg})}{\frac{2\pi}{c}d_{cg}d_{sa}f(y, d_{cg}) - d_{sa}\left(\frac{2\pi d_{cg}}{c} - \log\frac{2\pi r_g}{c}\right)} \quad . \quad . \quad . \quad . \quad (17)$$

Eqns. (16) and (17) give μ_a and μ_s as functions of the distance measured along the cathode surface, and determine the extent up to which the *inselbildung* effect is prominent in the characteristic curve of the tube. Since $\mu_a \gg \mu_s$ and beam power valves are usually operated above the knee of the plate characteristics, *inselbildung* would depend mainly on the variation of μ_s with y.

We note that when $d_{cg} \gg c$

$$f(y, d_{cg}) \simeq \tanh\frac{2\pi d_{cg}}{c} \simeq 1$$

and eqns. (16) and (17) become

$$\mu_a = \frac{d_{sa}d_{gs} - \frac{c}{2\pi}d_{ga}\log\frac{2\pi r_s}{c}}{\left(\frac{c}{2\pi}\right)^2\log\frac{2\pi r_s}{c}\log\frac{2\pi r_g}{c}} \quad . \quad . \quad . \quad (18)$$

$$\mu_s = \frac{d_{sa}d_{gs} - \frac{c}{2\pi}d_{ga}\log\frac{2\pi r_s}{c}}{-\frac{c}{2\pi}d_{sa}\log\frac{2\pi r_g}{c}} \quad . \quad . \quad . \quad (19)$$

Thus, when $d_{cg} \gg c$ the expressions for μ_a and μ_s become identical with those given earlier by Spangenberg.[3]

A concept useful in the study of electric field inside an electron tube is that of the capacitance coefficient, defined as the equivalent capacitance between two electrodes per unit area. If C_{cg}, C_{cs} and C_{ac} denote the capacitance coefficients between the cathode and the control grid, the screen grid and the cathode and the anode and the cathode respectively, the off-cathode potential gradient may be expressed as

$$\left(\frac{\delta V}{\delta x}\right)_{x=0} = 4\pi(C_{cg}V_g + C_{cs}V_s + C_{ac}V_a) \quad . \quad (20)$$

From eqns. (9), (10), (11), (15) and (20) we obtain readily

$$\left.\begin{aligned} C_{cg} &= \frac{1}{4\pi}\frac{L}{\Delta} \\ C_{cs} &= \frac{1}{4\pi}\frac{M}{\Delta} \\ C_{ac} &= \frac{1}{4\pi}\frac{N}{\Delta} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

We find from eqns. (15) and (21) that all the three capacitance coefficients may be expressed in the form

$$C = A + Bf(y, d_{cg}) \quad . \quad . \quad . \quad . \quad . \quad (22)$$

The capacitance coefficients therefore vary with y. We can define an average value of C by the relation

$$\bar{C} = \frac{2}{c}\int_0^{c/2} C dy$$

$$= A + \frac{2B}{c}\int_0^{c/2} f(y, d_{cg})dy \quad . \quad . \quad . \quad . \quad (23)$$

Putting $$Z = \coth\frac{2\pi d_{cg}}{c} \text{ and } \phi = \pi - \frac{2\pi y}{c}$$

$$\frac{2}{c}\int_0^{c/2} f(y, d_{cg})dy = \frac{1}{\pi}\int_0^{\pi}\frac{d\phi}{z + \sqrt{(z^2 - 1)}\phi}$$

$$= P_{-1}(z) = P_0(z) = 1$$

where $P_n(z)$ is the Legendre function of order n. One therefore obtains

$$\left.\begin{aligned} \overline{C_{cg}} &= \frac{2d_{ga}\log\frac{2\pi r_s}{c} - \frac{4\pi}{c}d_{sa}d_{gs}}{c\Delta} \\ \overline{C_{cs}} &= \frac{2d_{sa}\log\frac{2\pi r_g}{c}}{c\Delta} \\ \overline{C_{ac}} &= -\frac{\log\frac{2\pi r_s}{c}\log\frac{2\pi r_g}{c}}{\pi\Delta} \end{aligned}\right\} \quad . \quad . \quad . \quad (24)$$

The average values of the amplification factors are given by

$$\left.\begin{aligned} \overline{\mu_a} &= \frac{\overline{C_{cg}}}{\overline{C_{ac}}} \\ \overline{\mu_s} &= \frac{\overline{C_{cg}}}{\overline{C_{cs}}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

From eqns. (18), (19), (24) and (25) it is seen that Spangenberg's expressions may be regarded as those for the average values of the amplification factors.

It is easy to show that if d_{cs} and d_{sa} are both large compared to d_{cg}, eqns. (16) and (17) may be reduced to the form

$$\mu = \bar{\mu}\left[\frac{f(d_{cg}, y)}{\frac{2\pi d_{cg}}{c}\frac{f(d_{cg}, y) - 1}{\log\frac{2\pi r_g}{c}} + 1}\right]$$

where $\bar{\mu}$ is the average value of μ. The maximum value of μ occurs for $y = 0$. If this maximum value is written as μ_m, then the largest percentage deviation of the amplification factor from the mean value is obtained as

$$\frac{\mu_m - \bar{\mu}}{\bar{\mu}} \times 100 = \frac{(1 + \mu_1)\left(\coth\frac{\pi d_{cg}}{c} - 1\right)}{1 - \mu_1\left(\coth\frac{\pi d_{cg}}{c} - 1\right)} \times 100 \quad . \quad (26)$$

where $$\mu_1 = -\frac{2\pi d_{cg}}{c\log\frac{2\pi r_g}{c}}$$

Eqn. (26) has been plotted graphically in Fig. 2 with $2r_g/c$ as a parameter. It will be observed from the curves that the quantity $[(\mu_m - \bar{\mu})/\bar{\mu}] \times 100$ increases sharply as the ratio d_{cg}/c decreases below unity. This suggests that *inselbildung* would become appreciable when the cathode-grid distance became smaller than the grid pitch.

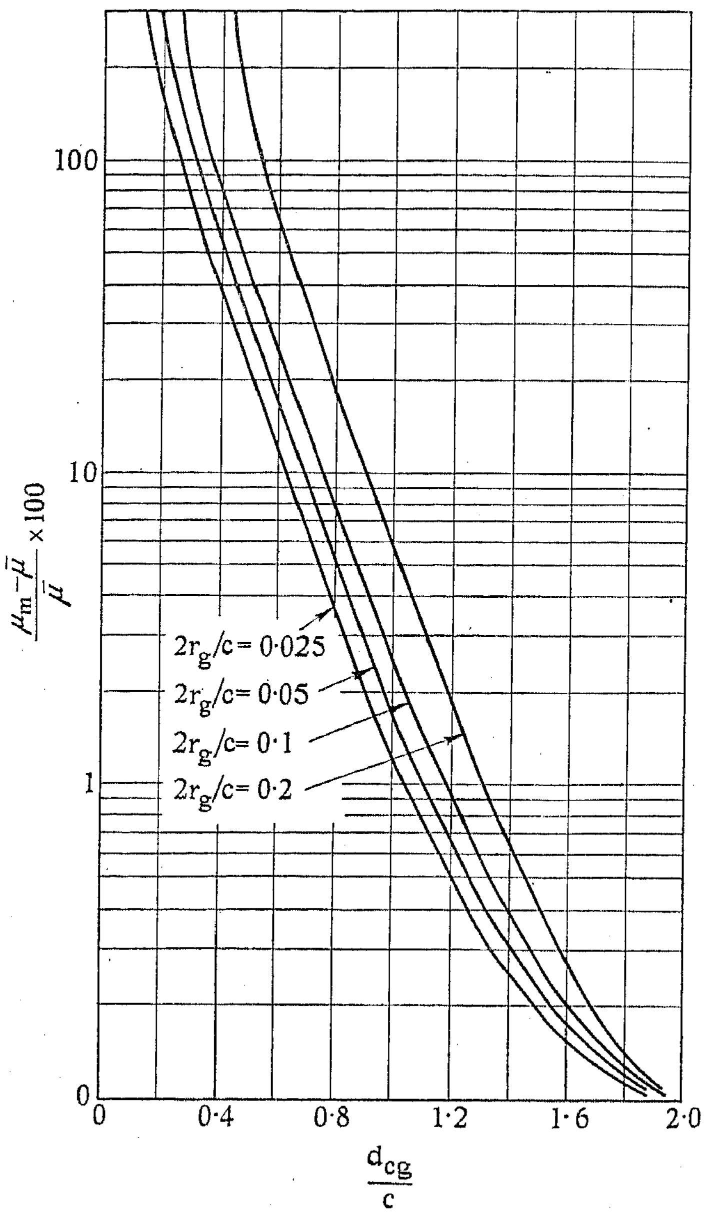


Fig. 2.—Illustrating how the largest percentage deviation of the amplification factor from the mean value ($\bar{\mu}$) varies with the cathode–control-grid distance measured in terms of the grid pitch.

(4) MUTUAL CONDUCTANCE OF A BEAM POWER VALVE

Anode current in a beam power valve, when no virtual cathode is formed between the screen grid and the anode, is given by

$$i_a = \frac{2{\cdot}336 \times 10^{-6} SDV_{eg}^{\alpha}}{d_{cg}^2} \quad . \quad . \quad . \quad . \quad (27)$$

where S = cathode area, $\alpha \simeq 1{\cdot}5$

$$D = \frac{\text{Anode current}}{\text{Total current}} \simeq 1 - \frac{2r_s}{c} \quad . \quad . \quad . \quad (28)$$

and

$$\left.\begin{aligned} V_{eg} &= \text{equivalent grid sheet potential} \\ &= \frac{V_g + \frac{V_s}{\mu_s} + \frac{V_a}{\mu_a}}{1 + \frac{1}{\mu_s} + \frac{1}{\mu_a} + \frac{4}{3}\frac{1}{\mu_c}} \\ \mu_c &= \text{amplification factor of grid with respect to cathode} \\ &= \frac{2\pi d_{cg}}{c\log\coth\frac{2\pi r_g}{c}\left(1 - \epsilon^{-\frac{4\pi d_{cg}}{c}}\right)} \\ &\text{for } d_{cg} < c \text{ and } 2r_g \ll c \end{aligned}\right\} \quad . \quad (29)$$

The factor 4/3 in the last term of the denominator takes into account the effect of space charge.[7] From eqn. (27),

$$g_m = 1{\cdot}5\left(\frac{2{\cdot}336 \times 10^{-6} SD}{d_{cg}^2}\right)^{2/3} i_a^{1/3}\ \text{amp/volt} \quad . \quad (30)$$

In computing the values of i_a and g_m with the help of eqns. (27) and (30), use should be made of the amplification factors as given by eqns. (16) and (17).

It is to be noted in this connection that the 3/2 power law of Child [eqn. (27)] is valid when the effect of the initial velocity

of the electrons is negligible. This is not, however, the case with closely spaced valves. For such valves the relation given by Langmuir and Fry is more accurate but is somewhat unwieldy for practical handling. Fortunately, however, in practical valves there are other factors such as contact difference of potential and end cooling which tend to mask the effect due to initial velocity. Because of this the results obtained from eqn. (27) may be taken to be substantially accurate for all practical purposes.[8] The masking effect due to contact potential difference is illustrated in Table 1 by considering the case of a

Table 1

Inter-electrode distance (i)	V (ii)	From Langmuir's law (iii)	From Langmuir's law with correction for contact p.d. (iv)	From Child's law (v)
mm	volts			
0·2	9	210	180	162
	3	59	38	26·8
	2	38	20	16·9
	1	20	7	6
0·1	9	800	700	650
	3	210	130	107·5
	2	130	75	67·5
	1	75	24	24

diode with an oxide cathode. For this the contact potential difference between the anode and the cathode has an average value of about −1 volt. In Table 1, columns (i), (ii) and (iii) give respectively the electrode dimension, the anode potential and the current as given by Langmuir and Fry's relations. Column (iv) gives the results of the preceding column corrected for contact potential difference, and column (v) the results as given by Child's relation. It is seen that the two sets of results given in columns (iv) and (v) are in satisfactory agreement.

(5) AN ILLUSTRATIVE EXAMPLE

We shall illustrate the accuracy and the usefulness of the results of analysis given in the foregoing Sections by considering the case of the Type 807 valve which had been used by an earlier worker for numerical computation.[2] The valve has the following dimensions:

$$d_{cg} = 0{\cdot}041\,\text{cm} \qquad \frac{2r_g}{c} = \frac{2r_s}{c} = 0{\cdot}100$$
$$d_{gs} = 0{\cdot}107\,\text{cm} \qquad c = 0{\cdot}078\,\text{cm}$$
$$d_{sa} = 0{\cdot}500\,\text{cm} \qquad s = 1{\cdot}820\,\text{cm}^2$$

Fig. 3 gives the variation of μ_a with y of the valve for one-half of a symmetrical section containing a single grid wire. The corresponding variation of μ_s is shown in Fig. 4. Variations in the other half of the section are just mirror images of those depicted by the curves in these Figures. It is seen from the

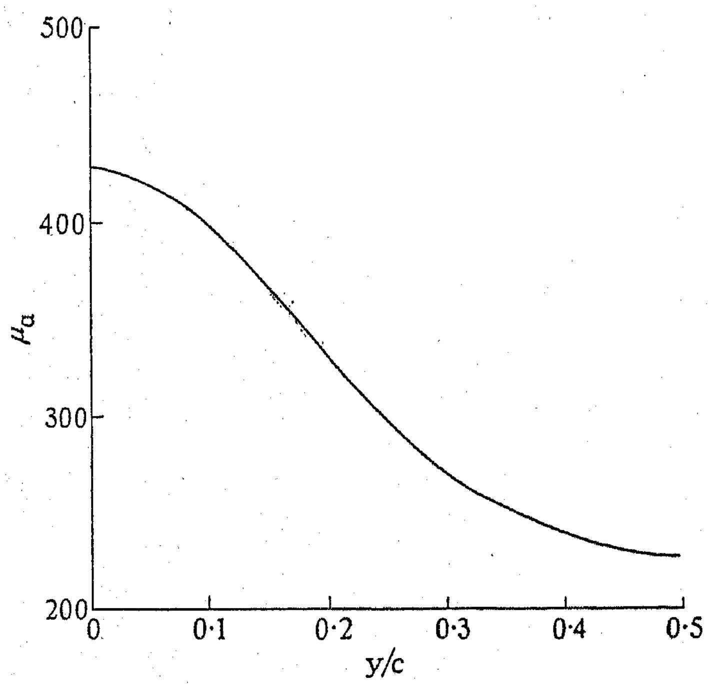


Fig. 3.—Illustrating how for a Type 807 beam power valve the amplification factor of the control grid with respect to the anode (μ_a) varies along the cathode surface.

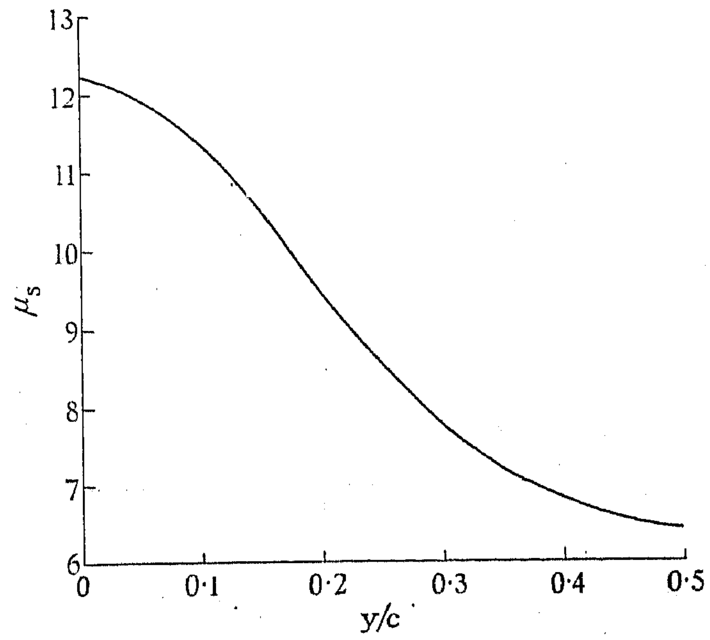


Fig. 4.—Illustrating how for a Type 807 beam power valve the amplification factor of the control grid with respect to the screen grid (μ_s) varies along the cathode surface.

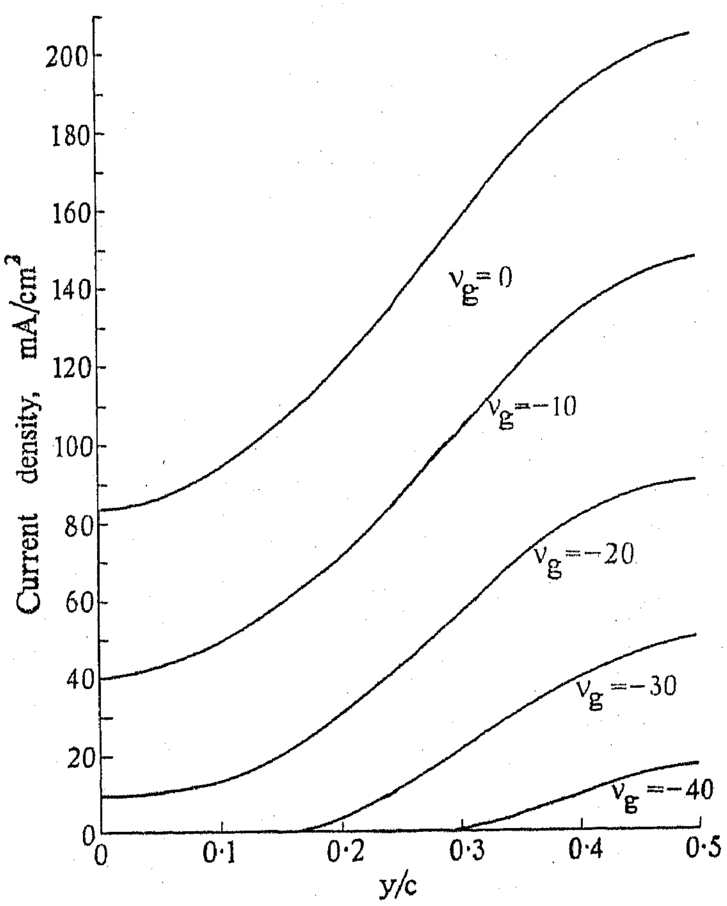


Fig. 5.—Illustrating how the space current density varies along the cathode surface of a Type 807 beam power valve for different values of V_g when V_a = 600 volts and V_s = 300 volts.

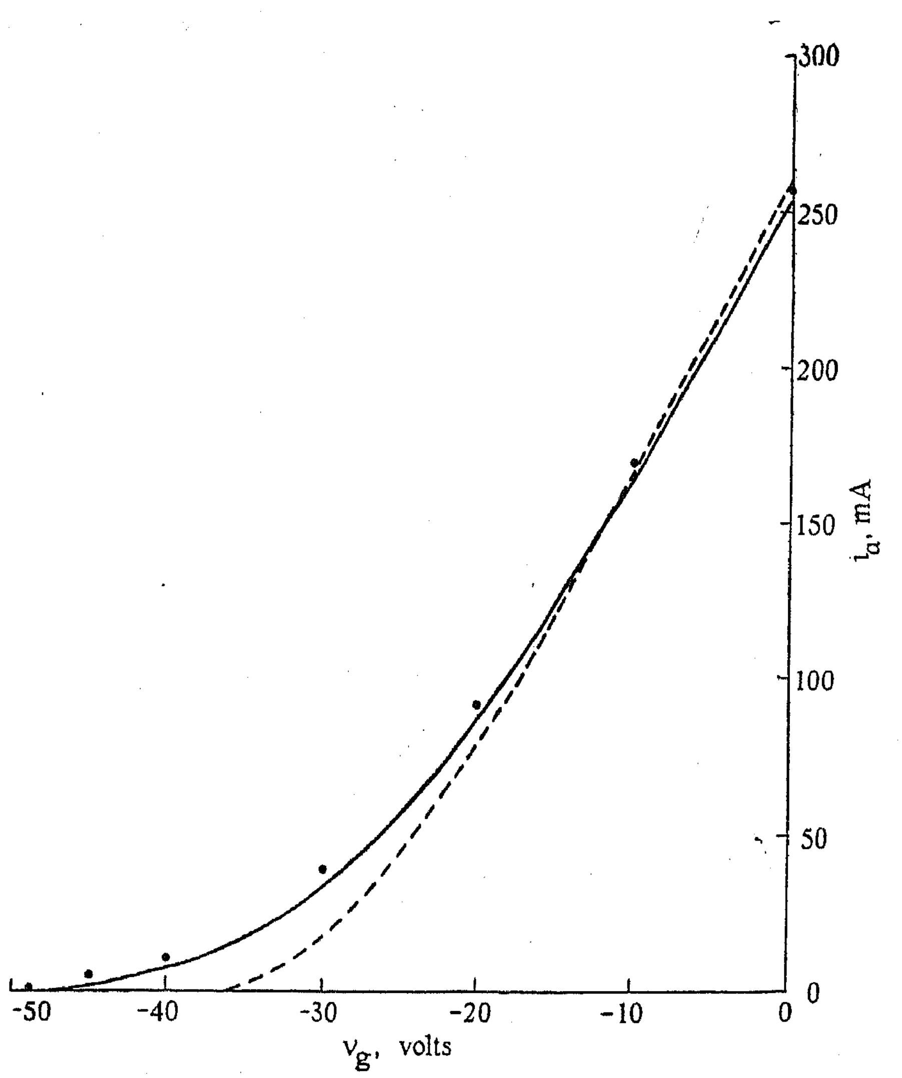


Fig. 6.—Mutual characteristic curves of Type 807 beam power valve, for $V_a = 600$ volts and $V_s = 300$ volts.

• • • Experimentally obtained values.
—— Curve obtained after present theoretical analysis.
---- Results as obtained from the earlier analysis of Wada.

curves that in the Type 807 valve under consideration the amplification factors vary by about 100%. This implies existence of appreciable *inselbildung*.

The above curves, together with eqn. (27), may also be used to examine the variation of space current density along the y-axis. This variation is shown by the curves in Fig. 5 for a number of values of V_g when $V_a = 600$ volts and $V_s = 300$ volts. Curves in Fig. 5 in their turn enable one to compute the mutual characteristic. This is shown by the continuous curve in Fig. 6. This curve may be compared, on the one hand, with the broken curve (in the same figure) which is plotted from the earlier computed results of Wada, and on the other, with the line joining the thick dots indicating points obtained experimentally. It is seen that the curve as obtained by the present analysis agrees better with the experimental results than the curve of Wada. In particular, the slope of the computed curve is in excellent agreement with that obtained experimentally. The theoretical value of g_m thus agrees almost exactly with the experimental value. For example, for $i_a \simeq 72$mA, we have $g_m \simeq 6\,000$ micromhos from the continuous curve (Fig. 6). This also is the value of g_m given in valve manuals. The slight lateral deviation of the theoretical curve from the experimental one might be due to contact difference of potential, initial emission velocity of the electrons and other such factors not taken into account.

(6) CONCLUDING REMARKS

The present investigation shows that in a beam power valve the *inselbildung* will be high if the control-grid pitch is larger than the cathode–control-grid distance. The analysis also shows that simple expressions for amplification constants may be derived by taking this effect into account. The expressions as derived remove to a large extent the discrepancy that exists at present between theoretically computed and experimentally obtained characteristics of the valves, even when the effects due to contact difference of potential, initial emission velocity and end cooling are neglected. Closer attention should therefore be given to the effect of possible "island formations" when designing such valves.

(7) ACKNOWLEDGMENTS

It is a pleasure to thank Professor S. K. Mitra for his constant encouragement and help in course of this work.

(8) REFERENCES

(1) SCHADE, O. H.: "Beam Power Tubes," *Proceedings of the Institute of Radio Engineers*, 1938, **26,** p. 137.
(2) WADA, M.: "Study on the Characteristics of a Beam Power Output Tube, Part I," *Scientific Reports of the Research Institute of Electrical Communication, Tohoku University*, 1952, **3**, p. 135.
(3) SPANGENBERG, K.: "Vacuum Tubes," McGraw-Hill Book Co., Inc., 1948.
(4) LIEBMANN, G.: "The Calculation of Amplifier Valve Characteristics," *Journal I.E.E.*, 1946, **93,** Part III, p. 138.
(5) KING, R. W.: "Calculation of Constants of a Thermionic Vacuum-Tube," *Physical Review*, 1920, **15**, p. 256.
(6) FREMLIN, J. H.: "Triode Amplification Factors," *Philosophical Magazine*, 1939, **27**, p. 709.
(7) TELLEGEN, B. D. H.: "The Effect of the Emission Current in a Triode," *Physica*, 1925, **5**, p. 301.
(8) HARRIS, I. A.: "On the Co-ordination of Circuit Requirements, Valve Characteristics and Electrode Design," *Journal of the British Institution of Radio Engineers*, 1949, **9**, p. 125.

Paper No. 1803 R
July 1955

621.376.2

CARRIER-LEAK IN RECTIFIER MODULATORS OF THE SHUNT TYPE

Factors affecting the Attainment of Very Low Leak Voltages and the Correlation of D.C. and A.C. Leak

By G. L. CONNON, B.Sc.(Eng.), Graduate.

(*The paper was first received* 4*th August, and in revised form* 26*th November*, 1954.)

SUMMARY

An experimental study has been undertaken in an endeavour to achieve an understanding of the mechanism of carrier-leak in shunt modulators, to elucidate some of the factors affecting the attainment of low leak levels (in the range 40 to 80 dB below the carrier voltage), and to investigate the possibilities of d.c. balancing of the rectifier bridge.

With a square-wave carrier voltage in a non-reactive circuit (or at low frequencies) conditions approximate to those of ideal switching; carrier-leak is a function of the unbalance of the rectifier resistances for two specific voltages (one positive and one negative) and is quite calculable. A very low total leak voltage (i.e. including all frequency components) can be obtained by means of a simple resistance balance. The correlation of d.c. and a.c. components of the leak is excellent if rectifiers with high or alternatively well-balanced back-resistances are used.

With a sinusoidal carrier supply, carrier-leak arises during the transition period of rectifier switching (i.e. during the change from forward to backward resistance and vice versa), and can therefore no longer be completely eliminated by a simple resistance control. Complete suppression of any one component (usually the fundamental) is theoretically possible in the non-reactive case, but in practice, unbalance of rectifier and stray reactances gives rise to a finite leak voltage. Results, even at low frequencies (less than 5 kc/s), are thus inferior to those obtained using a square-wave carrier, and in the experimental circuit it was difficult to obtain, by resistance balancing, fundamental components of the leak voltages lower than 0·1% of the carrier forward peak potential difference. Correlation between d.c. and a.c. balance conditions is also inferior, and zero d.c. leak might correspond to anything up to 0·4% leak of fundamental a.c. component. It is shown that the use of a capacitive as well as a resistive balance control can reduce leak voltages to a very low order, approaching—and with care equalling—zero on a short-term basis.

At higher frequencies, reactance unbalance plays a very large part in the production of carrier leak. It is demonstrated that the minimum fundamental leak voltage obtainable with a resistive balance alone rises linearly as frequency is increased, and that a combination of resistive and reactive balance controls can be used to achieve suppression of one component as good as, or better than, that obtained at low frequencies without a reactive balance.

The a.c./d.c. correlation at high frequencies is quite reasonable, even with a reactive unbalance. With an unbalance of 50 $\mu\mu$F at 1 kc/s (equivalent to the probable value of 0·5 $\mu\mu$F at 100 kc/s) correlation was such that for zero d.c. leak the a.c. fundamental leak did not exceed about 0·5% of the carrier forward peak potential difference.

It appears to be a fairly safe general conclusion that control of carrier-leak adjustment by the use of a d.c. meter is likely to prove very useful in practice and, provided that efficient rectifiers are used, should enable a.c. leak voltages to be maintained at least 40 dB below the carrier voltage across the rectifiers.

Under reactive conditions, the effect of carrier-generator impedance is complicated. If it is itself partly reactive, the capacitance balance becomes frequency dependent, i.e. a different adjustment of the balancing capacitor is required at each frequency. This dependence on the carrier-generator impedance is probably due to the fact that the rectifier bridge is not truly balanced over the whole switching cycle, but is merely adjusted to give a minimum output of one particular frequency.

Some results are given for the type of shunt modulator having no transformer—a useful alternative to the Cowan type. Very low leak voltages are shown to be obtainable by adjustment of the relative value of the two coupling capacitors.

(1) INTRODUCTION

Amplitude modulation is basically a multiplicative process, and the output can be expressed as the product of the input signal and a time function—termed the modulating function—which is dependent on the carrier input. Amplitude modulators can be broadly divided into two classes: those where the modulating function is dependent on the instantaneous amplitude of the carrier (multi-electrode valve modulators, etc.), and those where the modulating function is dependent only on the polarity of the carrier (switching modulators). The shunt (or Cowan) modulator is a single-balanced modulator of the latter type—single-balanced because the output ideally contains no component at carrier frequency.

The basic circuit of the shunt modulator is shown in Fig. 1,

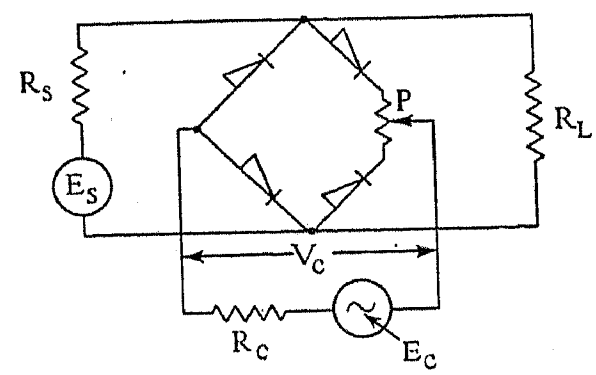


Fig. 1.—Basic shunt modulator.

E_S = Signal source.
E_C = Carrier source.
R_L = Load.
P = Carrier-balance potentiometer.

and the action of the circuit is as follows. Assuming that the carrier voltage developed across each rectifier is much larger than the signal voltage, the rectifier bridge shunting the load R_L alternates between a high and a low impedance, compared with R_L, according to the polarity of the carrier voltage. The waveform of the output across R_L is therefore a chopped version of the input signal. In the ideal case, with a sinusoidal input signal, this output can be shown to consist of a component at signal frequency, f_s, plus an infinite series of frequency components of the form $nf_c \pm f_s$ where f_c is the carrier frequency and n is odd. In this ideal case there is no component at f_c present in the output, the rectifier bridge being perfectly balanced. This is not, of course, possible in a practical circuit, and the unbalance component in the output at carrier frequency is termed carrier leak.

Rectifier modulators are most commonly used, in conjunction with a filter, as frequency-changers, and in many cases a large dynamic range of signal input is desirable. The upper limit of

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Connon is at Her Majesty's Underwater Detection Establishment.

permissible input level for a given carrier voltage is set by overloading and consequent non-linearity of the modulator, the overloading occurring when the signal voltage developed across the rectifiers in opposition to the carrier voltage becomes sufficiently large to prevent complete switching of the rectifiers. The lower limit of input signal is set by carrier leak. It is therefore desirable to reduce the latter as much as possible.

A certain degree of carrier-leak balance can be obtained in any nominally balanced rectifier modulator by selection of rectifiers or the use of a balancing potentiometer, and the calculation of the effect of unbalances in an idealized non-reactive modulator with a square-wave carrier is quite simple. A very different situation exists, however, when low leak-levels are required in real modulators with imperfect switching in partly reactive circuits. The subject of carrier leak has received little attention in published work, the only treatment known to the author going beyond the simplest conception of the problem being a study by D. G. Tucker[1,5] of the effect of carrier-generator resistance on carrier leak in ring and shunt modulators using a sinusoidal carrier, rectifiers with a slow transition from high to low resistance and vice versa, and perfectly non-reactive circuit elements. Another important contribution from the practical point of view is a note[2] by the same author showing that in single-balanced modulators, such as the shunt modulator, there may be, with suitable design, a close correlation between the d.c. and a.c. components of the carrier leak; this would permit maintenance and control of carrier leak to be effected by a d.c. meter instead of the usual expensive frequency-selective valve voltmeter. The analysis given was limited by the assumption of a square-wave carrier and non-reactive circuit elements.

It is clear that a much more thorough study of carrier-leak problems is required if leak voltages of less than, say, 0·1% of the carrier voltage across the rectifiers are to be attained. A mathematical approach to the study appeared much too formidable to be of much practical value, and therefore an experimental study was undertaken to elucidate some of the factors affecting the attainment of very low leak-levels (in the range 40 to 80 dB below the carrier voltage), with particular reference to the correlation between d.c. and a.c. components of the carrier leak.

The majority of the experiments were carried out with two particular modulator circuits, using thermionic diodes with a forward resistance of about 300 ohms, carrier forward peak potential differences of about 1 volt, and a modulator load resistance of the order of 3 000 ohms. It is thought that such a modulator is reasonably typical of many used in practice, and that it is therefore worth while to give some detailed experimental results in the succeeding parts of the paper.

(2) CORRELATION BETWEEN A.C. AND D.C. CARRIER-LEAK VOLTAGES WITH SQUARE-WAVE SWITCHING (THEORETICAL)

For convenience, the contents of the note previously referred to[2] are summarized here.

It is suggested that for shunt (Cowan) modulators employing a square-wave switching (carrier) voltage, a carrier-leak balance sufficiently accurate for most practical purposes can be obtained using a d.c. meter to balance the d.c. component of the carrier-leak current, as opposed to an a.c. voltage measurement with a wave analyser or calibrated receiver. Monitoring and maintenance of modulators in service is thereby greatly simplified.

Fig. 1 shows a rectifier modulator of the shunt type, and the following assumptions are made:

(a) Square-wave switching is employed.
(b) The circuit is non-reactive.
(c) The rectifiers are ideally switched, i.e. they switch instantaneously at zero voltage from a constant forward resistance (r_f) to a constant backward resistance (r_b).

Owing to the presence of a finite carrier-source resistance R_c (Fig. 1), V_c will be as in Fig. 2(b). The subscripts f and b refer to forward and backward half-cycles respectively. The general form of the carrier-leak voltage across the load R is shown in

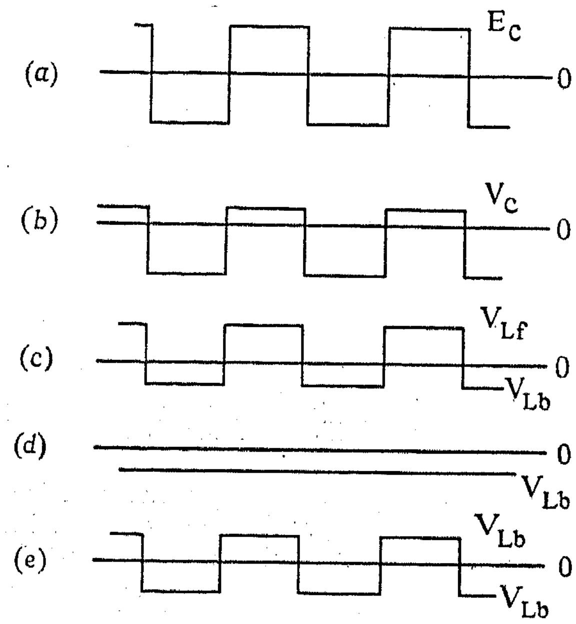


Fig. 2.—Waveforms of carrier-leak voltage of shunt modulator with square-wave carrier supply.

(a) Carrier e.m.f.
(b) Carrier potential difference (reduction in amplitude of positive half-cycles if $R_c \neq 0$).
(c) Carrier-leak voltage across load (unbalanced).
(d) Carrier-leak voltage across load (a.c.-balanced).
(e) Carrier-leak voltage across load (d.c.-balanced).

Fig. 2(c). The leak voltage during the forward half-cycle (V_{Lf}) can be varied over a range of positive or negative values by means of the balancing potentiometer P (usually of the order of a few hundred ohms). The backward leak voltage (V_{Lb}) will be very small if r_b is large, and will be virtually unaffected by variation of P.

The two methods of balancing give a carrier-leak waveform as shown in Figs. 2(d) and 2(e).

Consider these separately:

A.C. Balance [*Fig.* 2(*d*)].

A.C. balance is obtained by adjusting P for minimum a.c. leak at carrier frequency (measured on wave analyser). The residual leak is a direct voltage of amplitude V_{Lb}.

D.C. Balance [*Fig.* 2(*e*)].

Here P is adjusted so that the mean d.c. level of the leak waveform, as measured with a microammeter in series with the load, is zero. This means that V_{Lf} is made equal but opposite in sign to V_{Lb}, and the residual fundamental leak-voltage component is $4/\pi V_{Lb}$ (peak).

If r_b is high (say 100 kilohms or more), then V_{Lb} will be very small, and the residual a.c. leak for the d.c. balanced condition will, for most purposes, be negligible. If r_b were infinite, then V_{Lb} would be zero and therefore a.c. and d.c. balance conditions would exactly correspond. This condition is closely approximated in modulators using thermionic diodes, for which r_b is almost infinite, as switching elements.

(3) SQUARE-WAVE CARRIER AT LOW FREQUENCIES

The above theory was confirmed by measurements on a shunt modulator (Fig. 3) using diode rectifiers each shunted by a 1-megohm resistor (to give a finite and known back resistance). The carrier generator was isolated from the modulator by means of a transformer, thus enabling it and the modulator load to

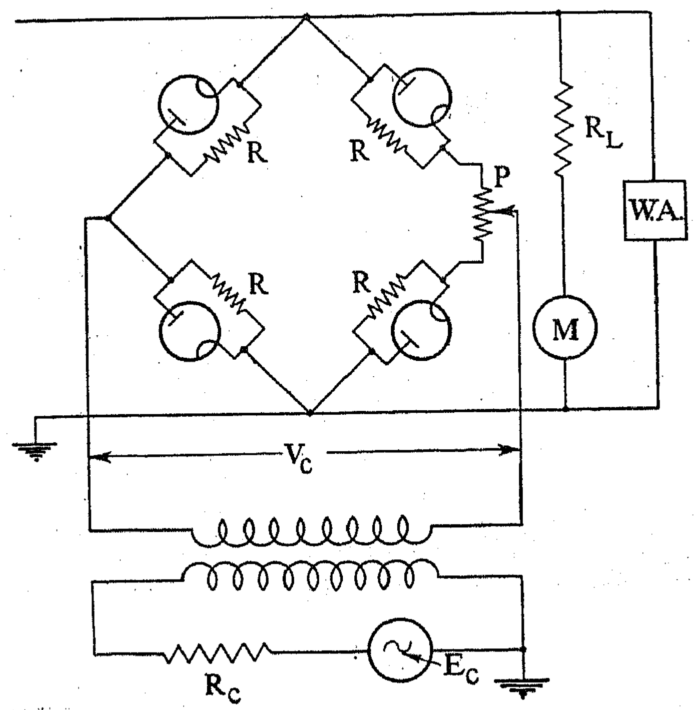


Fig. 3.—Circuit used for experimental work on carrier-leak.

R = 1 megohm. P = 500 ohms.
M = Unipot centre-zero microammeter.
WA = 4 c/s-bandwidth wave analyser.
E_C = Carrier e.m.f.
V_C = Carrier potential difference.
Transformer ratio 1 : 1.
Rectifiers CV140.

have one side earthed. The square-wave switching waveform was obtained from a cathode-coupled trigger pair, triggered by the zero-crossings of the output from a beat-frequency oscillator. A centre-zero unipivot microammeter was used to measure the d.c. component of the carrier-leak current in the load.

Experimental results (not given in detail) confirmed the theory of the preceding Section. At low frequencies (1 kc/s or so), where the modulator can be assumed to be non-reactive, zero d.c. leak corresponded very closely to a minimum a.c. leak, the value of the latter being of the order of 0·1% of the carrier forward peak potential difference.*

(4) SINE-WAVE CARRIER AT LOW FREQUENCIES

(4.1) Basic Theory

When a sinsusoidal switching voltage is used, the simple approach of Section 2 is no longer valid, since the major part of the carrier leak is due to unbalance during the transition period, i.e. while the rectifiers are changing from forward to backward resistance values and vice versa. For detailed examination of carrier leak, therefore, the assumption of ideal switching of rectifiers must be abandoned.

The impedance characteristic of a thermionic diode follows very closely an exponential law:[3,5]

$$r = r_f + k\varepsilon^{-qV} \text{ (see Fig. 4)}$$

where r is the a.c. resistance of the diode (defined as dV/dI), and r_f, k and q are constants. It has been shown experimentally that the parameter q is fairly constant for a given type of diode, while r_f and k vary from sample to sample, and with heater current.

Thus the simple resistance balance P cannot completely eliminate carrier leak, since adjustment of P can only compensate for differences in r_f. The residual leak is then due to differences in the value of k.

* Unless otherwise stated, all results quoted in the paper refer to a shunt modulator as shown in Fig. 3.
Carrier source resistance (R_c) ≑ 2·2 kilohms.
Carrier potential difference (V_c) = 1 volt forward peak.
Modulator load (R) = 3·3 kilohms.
Leak voltages are quoted at fundamental frequency.

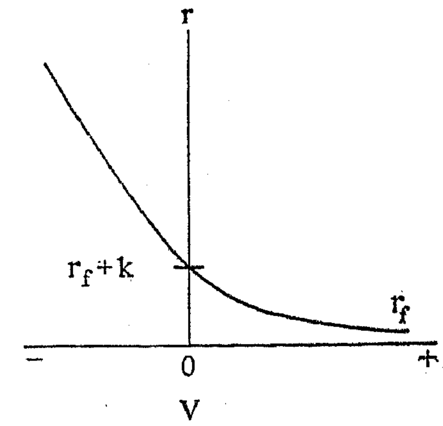


Fig. 4.—Typical diode a.c. resistance characteristic.

Represented closely by the equation:
$r = r_f + k\varepsilon^{-qV}$
(Logarithmic scale for r.)

Fig. 5 shows the form of carrier leak to be expected. The waveform can be divided as follows:

Period A.—Diode resistance is very high (r_b) and leak voltage therefore negligible.

Period B.—Transition from r_b to r_f takes place; leak voltage is due to difference in value of k for the diodes.

Period C.—Diodes are biased to forward resistance, balanced by means of P. Leak voltage can be made zero.

Period D.—Transition from r_f to r_b.

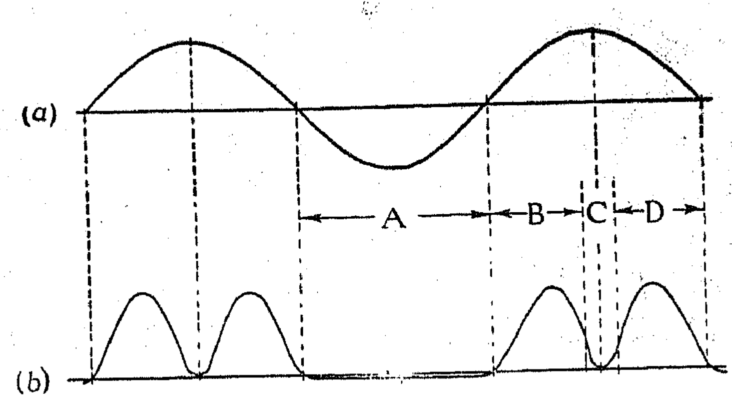


Fig. 5.—Form of carrier-leak voltage with sinusoidal carrier voltage.

Rectifiers conducting on positive half-cycle.

Period A: Leak voltage low owing to high back resistance of rectifiers.
Period B: High leak-voltage due to unbalance of rectifiers during transition from non-conducting to fully conducting condition.
Period C: Leak voltage reduced by means of balancing potentiometer P (Fig. 1).
Period D: Transition of rectifiers from fully conducting to non-conducting condition.

The complete waveform (i.e. as in Fig. 5) can be calculated using the exponential law given above to represent the rectifier characteristics. The waveform obtained in practice was in close agreement with this, and is illustrated in Fig. 6(*d*), which shows the carrier leak obtained at 3 kc/s from a shunt modulator as in Fig. 3, using a sinusoidal carrier e.m.f.

Variation of P controls the magnitude of the leak voltage during the period C, by balancing the forward resistances of the diodes. This voltage can be made positive or negative if the range of P is sufficiently large. The backward leak (period A) is virtually unaffected. This is illustrated by Fig. 6 showing the effect of variation of P on the carrier-leak waveform of the modulator of Fig. 3 at 3 kc/s.

(4.2) Balance of One Frequency Component

In the case of a perfectly non-reactive circuit, a perfect balance of any one frequency-component of the carrier-leak voltage (the fundamental, say) appears to be possible using only a resistive adjustment, in spite of the time-varying nature of the rectifier resistances. This follows since these non-linear components can be replaced, at fundamental frequency and for a given funda-

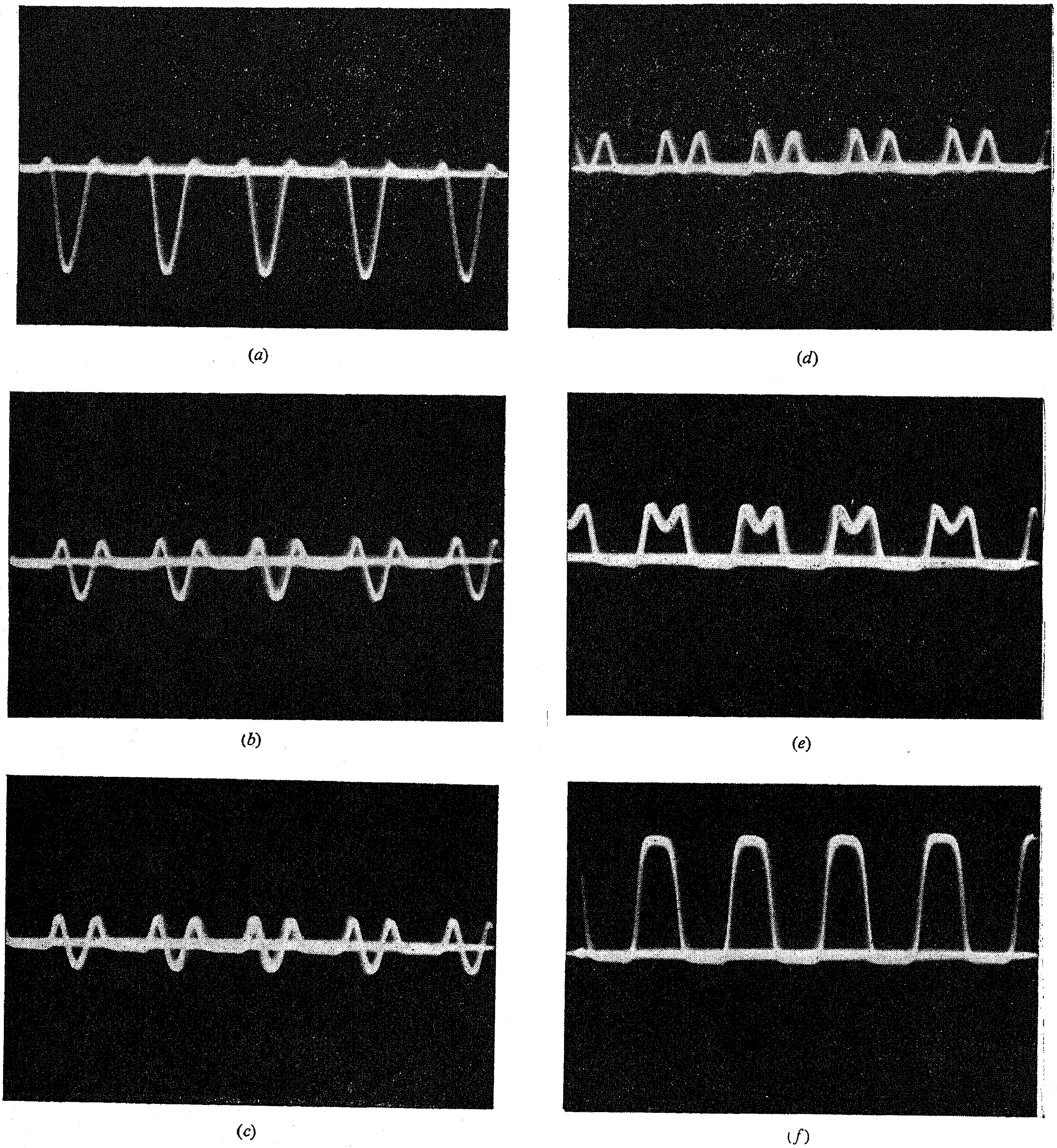


Fig. 6.—Carrier-leak voltage waveforms illustrating effect of balancing potentiometer P.

Circuit as in Fig. 3, with sinusoidal carrier supply.
f_c = 3 kc/s, R_G = 500 ohms, R_L = 3·3 kilohms, and V_G = 1 volt forward peak.
Reference line at earth potential.
D.C. leak and r.m.s. fundamental a.c. leak voltages:

	μA	mV
(*a*)	−4	15
(*b*)	0	2·5 (d.c.-balanced)
(*c*)	0·8	0·2 (a.c.-balanced)
(*d*)	2	3·5
(*e*)	4	10
(*f*)	9	25

mental voltage across them, by their fundamental resistances—hypothetical values which are voltage dependent, but not time-varying.* This technique has been described fully by Moullin[6] and Slemon,[7] although not in connection with modulators.

A recent theoretical analysis of carrier leak in modulators by Tucker[9] also confirms this, although the result obtained is not perfectly general.

(4.3) Correlation between A.C. and D.C. Balance

The values of residual a.c. leak (fundamental component) and mean current through the load are given in Fig. 6 for various settings of P. It is always possible to adjust P so that the mean current through the load is zero (i.e. d.c. balanced condition) since P affects only the forward leak. This will clearly occur when the areas enclosed by the leak waveform above and below the zero axis are equal. [See Fig. 6(*b*).]

This, of course, does not necessarily correspond to minimum fundamental leak, and the residual a.c. leak in Fig. 6(*b*) was 2·5 mV. Minimum a.c. leak (200 μV) in fact corresponded to a mean current of 0·8 μA.

This result is fairly typical for the type of rectifier used, and it will be noted that although it can be further reduced, the a.c. leak obtained at low frequencies by balancing P to give zero mean current through the load is sufficiently small for most applications. (With a carrier potential difference of 1·0 volt forward peak, a fundamental leak voltage of less than 10 mV is often satisfactory.)

(4.4) Production of Higher Harmonics

This paper is mainly concerned with the fundamental component of carrier-leak voltage, which, as stated in Section 4.2, can theoretically be eliminated in the non-reactive case. It should be noted, however, that the carrier-leak voltage will contain higher harmonics of the carrier frequency, the amplitude of which may be much higher than the fundamental when the modulator is balanced for fundamental. For example, in a case where the fundamental (4 kc/s) component of the carrier leak was 0·5 mV, the second and third harmonics were found to be 20 dB higher than this, and the fourth harmonic 6 dB higher. The leak voltage contains both odd and even harmonics, as opposed to the case of a square-wave carrier, where only odd harmonics will be present. Another case revealed harmonics up to the tenth having amplitudes from 6 to 30 dB greater than the fundamental (limitations of the wave analyser prevented investigation of higher harmonics). Any of these components can be greatly reduced by readjustment of P, but doing so of course increases the fundamental. In this case the sixth harmonic was reduced to about 0·5 mV, and the fundamental component rose from zero to over 30 mV.

These higher harmonics are not as a rule troublesome, since they can readily be removed by filtering at the modulator output. However, in some applications, e.g. where the switching frequency is lower than the signal frequency, they can be of great importance.

(4.5) The Effect of Carrier-Generator Resistance (R_c)

The relation between the fundamental component of the a.c. leak voltage and the mean direct current is shown in Fig. 7 for various values of R_c. The effect of R_c within the range taken (15 ohms to 10 kilohms) on the correlation between d.c. and a.c. leaks, is clearly small. The frequency in all cases was 3 kc/s, and carrier potential difference 1·0 volt forward peak. It is seen that the a.c. leak is less than 10 mV over a range of mean current of 6 μA.

* The fundamental resistance of a non-linear impedance is defined as the ratio of the fundamental r.m.s. voltage across it to the fundamental r.m.s. current flowing through it.

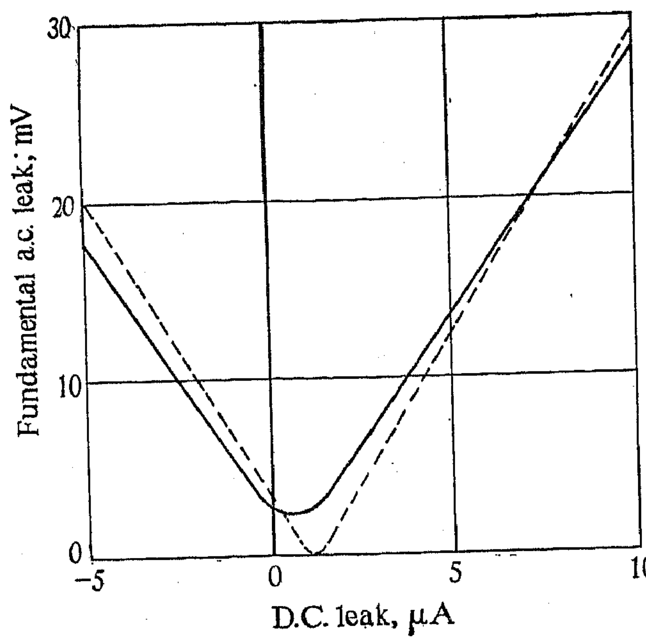


Fig. 7.—Effect of carrier-generator resistance (R_C) on carrier-leak voltage.

Circuit as in Fig. 3.
$f = 3$ kc/s, $V_C = 1$ volt forward peak, $R_L = 3·3$ kilohms.
———— $R_C = 10$ kilohms.
- - - - - - $R_C = 15$ ohms.
($R_C = 2·2$ kilohms and $R_C = 500$ ohms gave intermediate results.)

The higher value of minimum leak for $R_c = 10$ kilohms is thought to be due to the increased effect of reactance. This will be discussed in greater detail in Section 5.

(4.6) Effect of Variation of Carrier Voltage

It is often assumed that with a sufficiently high carrier voltage, the rectifiers switch from a constant forward resistance, r_f, to a large backward resistance, r_b, and that when balanced, therefore, the a.c. leak should be directly proportional to the carrier voltage —i.e. the balance is not affected by variation of the latter. This is not, in fact, the case, and balance of carrier leak must vary with carrier voltage (excluding the purely academic cases of four ideal rectifiers, four identical rectifiers, or square-wave carrier and constant values for r_f and r_b).*

This follows since:

(*a*) The forward resistance r_f is never absolutely constant, but falls slightly with increasing voltage across the rectifier.

(*b*) With a sinusoidal carrier, the finite transition period of the rectifiers causes dependence on carrier voltage, independent of the constancy or otherwise of r_f. That this must be so becomes obvious if the rectifiers are replaced by their fundamental resistances (see Section 4.2), the values of which are voltage dependent, the factor of proportionality not necessarily being the same for each rectifier.

The effect of carrier-voltage variation is not serious unless very low leak voltages are required.

This is illustrated in Fig. 8, which shows the change in a.c. leak over a range of carrier e.m.f. of $\pm 50\%$ of the voltage at which the modulator was balanced (in this case 1·0 volt forward peak across the rectifiers, corresponding to 2·6 volts r.m.s. carrier e.m.f.). Curve (*a*) shows the variation for an a.c. balance at the nominal voltage, curve (*b*) for a d.c. balance. It will be noted that over this wide range the a.c. leak did not exceed 5 mV. The leak resulting from an original d.c. balance actually decreased in one direction as it approached the a.c. balance condition at a higher carrier voltage. This can, of course, occur on either side of the nominal voltage, depending upon the diode characteristics.

* An Appendix to Reference 9 shows theoretically that in a simple shunt modulator with r_b high, carrier leak is zero if all four rectifiers are identical, but is proportional to high powers of V_c if there is an initial unbalance.

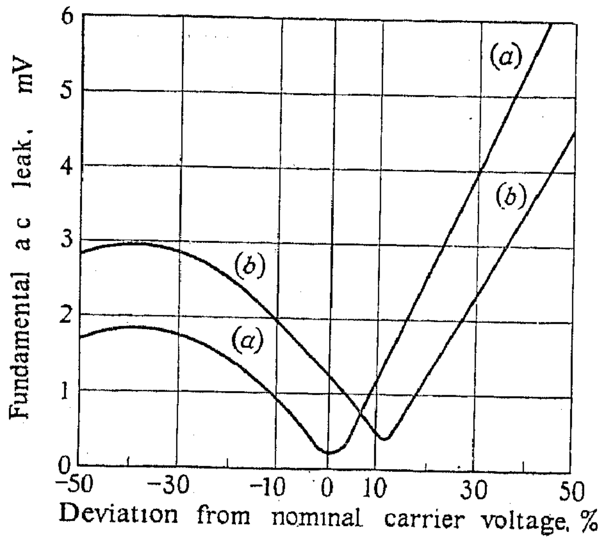


Fig. 8.—Effect of carrier voltage (V_C) on a.c. leak.

Circuit as in Fig. 3.
$f = 3$ kc/s, $R_L = 3{\cdot}3$ kilohms and $R_O = 2{\cdot}2$ kilohms. Nominal $V_O = 1$ volt forward peak.

(*a*) Modulator a.c.-balanced.
(*b*) Modulator d.c.-balanced.

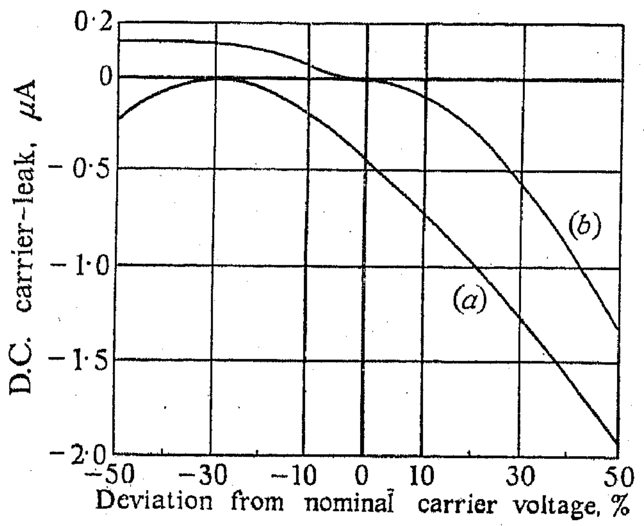


Fig. 9.—Effect of carrier voltage (V_C) on d.c. leak.

Conditions as for Fig. 8.
(*a*) Modulator a.c.-balanced.
(*b*) Modulator d.c.-balanced.

The variation of the d.c. leak (mean current through load) is shown in Fig. 9.

(4.7) Effect of Back Resistance

As stated earlier, the carrier-leak voltage during the backward half-cycle (diodes non-conducting) is negligibly small, since the load resistance is very low compared with that of the rectifier network. The effect of unbalance of the rectifier backward resistance (in this instance, of the resistors shunting the diodes—see Fig. 3) is therefore small, if rectifiers with a very high back-resistance are used. With 1 volt forward peak carrier potential-difference at 10 kc/s applied to the circuit of Fig. 3, a fundamental a.c. leak voltage (a.c. balanced) of $0{\cdot}2$ mV with each shunting resistor equal to 1 megohm $\pm 5\%$ rose to $2{\cdot}2$ mV when one resistor was reduced to 500 kilohms. The d.c. leak changed only from $0{\cdot}9$ to $0{\cdot}8\,\mu$A.

(4.8) Effect of Input Signal

The paper is entirely concerned with carrier leak measured or calculated in the absence of signal. The presence of voltage components due to a signal applied at the input terminals of the modulator can certainly modify the balance of the bridge in a practical case, but the resulting increase in carrier leak is very difficult to predict. It has been shown recently,[9] however, that the application of a signal affects carrier leak only if a carrier unbalance exists in the absence of signal, and that a perfectly balanced modulator should not be unbalanced by the signal voltage. Measurements have shown that with either square-wave or sinusoidal carrier voltages, the effect of signal on carrier leak is very slight, provided the amplitude of the former does not approach overload level.

(5) EFFECT OF REACTANCE

(5.1) Effect at Low Frequency

The foregoing Sections have applied to the case of a non-reactive circuit obtained in practice by working at low frequencies. At higher frequencies, the effect of reactive unbalance due to rectifier and stray capacitance becomes of increased importance. Unfortunately, exact mathematical analysis of modulators with reactive components is difficult and in many cases impossible.

One effect of reactance unbalance is an increase with frequency of the minimum a.c. leak obtainable. This is illustrated by Fig. 10(*a*), which is representative of the variation of minimum

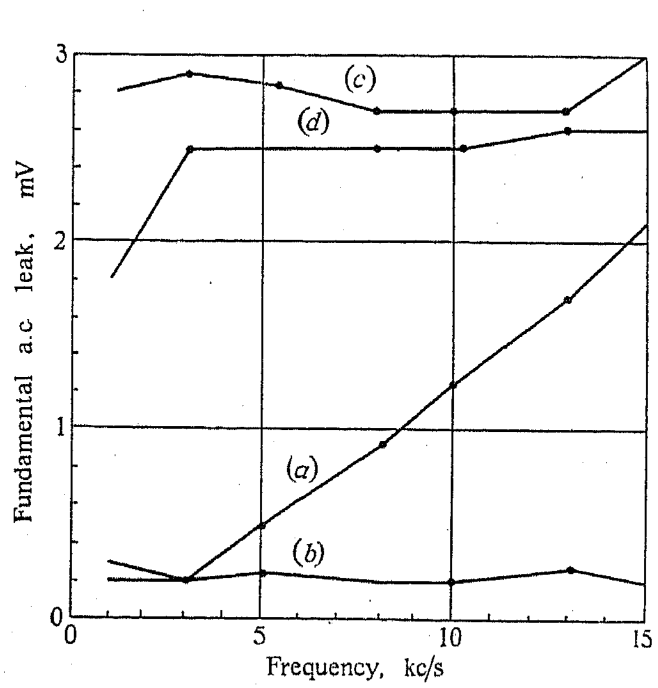


Fig. 10.—Effect of reactive unbalance on a.c. carrier leak.

Measured with circuit of Fig. 11.
(*a*) Modulator a.c.-balanced—no reactive balance.
(*b*) Modulator a.c.-balanced—with reactive balance.
(*c*) Modulator d.c.-balanced—no reactive balance.
(*d*) Modulator d.c.-balanced—with reactive balance.

a.c. leak with frequency. The leak voltage was reduced greatly by balancing the stray capacitance of the modulator* as in Fig. 11, where a small variable capacitor can be connected across whichever arm is necessary. The use of a screened and balanced isolating transformer in the carrier path minimized the effect of winding capacitances by balancing them to earth.

This arrangement reduced the carrier leak at frequencies up to 15 kc/s to a constant small value (approximately $0{\cdot}2$ mV) as in Fig. 10(*b*). Note, however, that to obtain the flat response shown it was necessary to re-trim the balancing capacitor at each frequency. This is discussed in Section 5.3.

Fig. 10(*c*) and 10(*d*) show similar curves with the modulator d.c.-balanced at each frequency. The a.c. leak without reactive balance, although higher, is more nearly constant than in the a.c.-balance case.

(5.2) Effect at Higher Frequencies

The degree of reactive unbalance can be increased, and the effect at higher frequencies thereby simulated, by the addition of

* By the same reasoning as in Section 4.2, the fundamental component (or any one component) can theoretically be completely eliminated by the use of a reactive and a resistive balance.

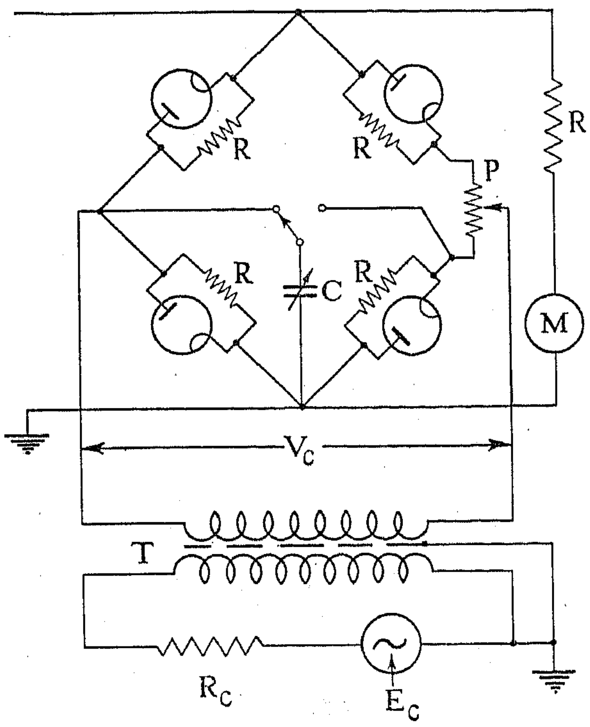


Fig. 11.—Shunt modulator with reactive balance.
V_O = 1 volt forward peak. R = 1 megohm. C = 0–100 $\mu\mu$F.
R_O = 2·2 kilohms. R_L = 3·3 kilohms. P = 500 ohms.
M = Centre-zero microammeter.
T = 1 : 1 screened and balanced transformer.

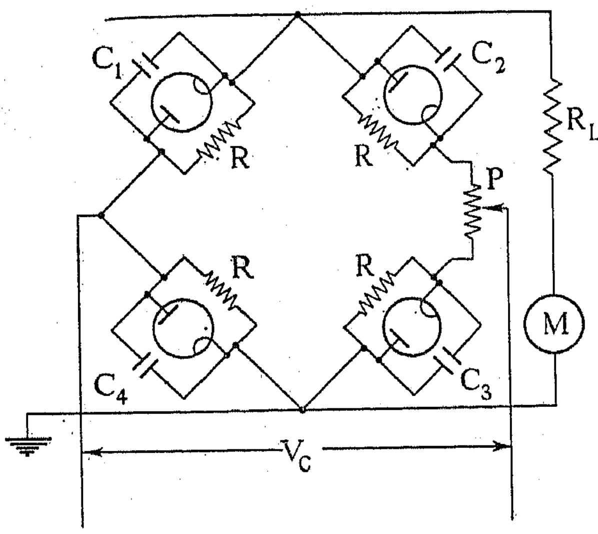


Fig. 12.—Modulator used to simulate effects at higher frequencies.
R_L = 2·2 kilohms, R = 1 megohm, P = 500 ohms.
$C_1 = C_2 = C_3 = C_4$ = 390 $\mu\mu$F ± 20%.
Diodes: CV140.
M = Centre-zero microammeter.

external capacitors across the rectifier elements, as in Fig. 12. The effect of reactance can then be studied in greater detail, and an estimate can be made of the modulator performance at higher frequencies. The capacitance chosen was 390 $\mu\mu$F, being approximately equivalent at 1 kc/s to 4 $\mu\mu$F at 100 kc/s, a probable practical value.

The effect on the waveform of adding these capacitors is illustrated in Fig. 13, which shows the effect at 1 kc/s of adding 390 $\mu\mu$F capacitors chosen at random across each arm in turn. It will be noted that the addition of 390 $\mu\mu$F (nominal) across all four arms returned the modulator almost to the original condition, as would be expected. The fact that the balance was actually improved is, of course, coincidental. The worst condition (with C_1 alone) gave a leak of 5·5 mV.

The accuracy of balance required, of course, increases with frequency. In another experiment, with 390 $\mu\mu$F (nominal) across each arm, the leak at 1 kc/s was as in Fig. 14(*a*); the fundamental component of a.c. leak due to a small reactance misbalance was 0·05 mV. This can be reduced to zero if the capacitors are balanced, as shown later. Fig. 14(*b*) shows the waveform obtained with C_1 removed; the leak rose to 2 mV. When C_1 was replaced and the frequency increased to 5 kc/s [Fig. 14(*c*)], the leak rose to 1·3 mV, and its waveform became similar to the unbalanced case of Fig. 14(*b*). That the increased leak at the higher frequency is due to reactance unbalance was further demonstrated by adjusting one of the capacitive arms, the effect of which was to return the waveform to the low-frequency form and to reduce the a.c. leak very nearly to zero [Fig. 14(*d*)].

It appears, therefore, that the major part of the carrier leak at high frequency is due to reactance unbalance, and that the leak could be reduced to a very low value by a reactive balance (by means of a variable capacitor across one arm) as well as a resistive balance (by means of P). There is a frequency-dependent effect, however, such that a reactive balance at one frequency does not hold for a different frequency.

(5.3) Frequency-Dependent Effect

With the circuit used for the above experiments, it was found that the capacitance required to balance the modulator varied to some extent with frequency. This was found to be a function of either the direct or the unbalance impedance in the carrier-generator circuit, which includes a transformer and undoubtedly varies with frequency. It is quite understandable that the reactive nature of the carrier-generator impedance might affect the leak, and since the balance condition obtained is only a compromise to give a minimum of one particular frequency-component and is not an actual balance at all points in time, the whole effect is not unreasonable. Nevertheless, tests to determine the exact nature of the causes of the effect have not yet proved conclusive.

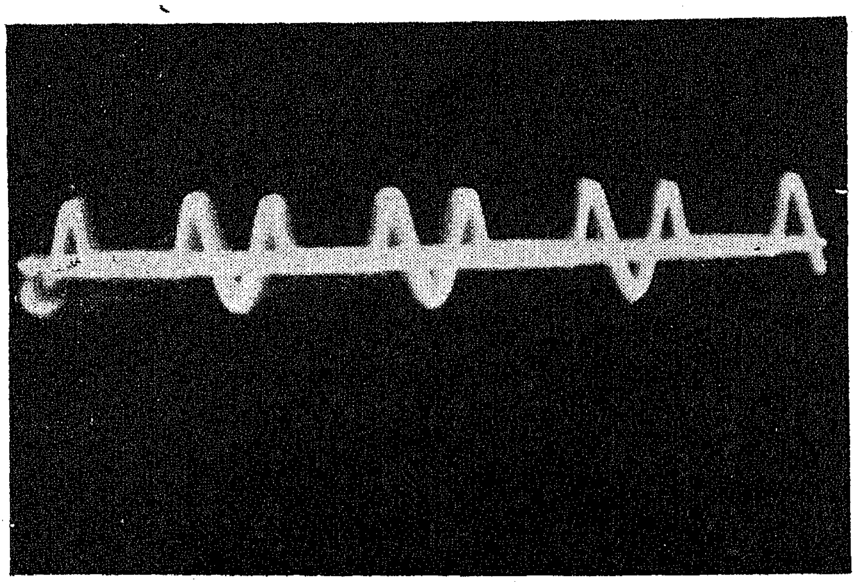
(5.3.1) Modulator A.C.-Balanced.

To enable a reactance balance to be made with greater accuracy, the circuit of Fig. 12 was used with C_2 made up of a fixed capacitor in parallel with a standard air condenser, variable from 0 to 120 $\mu\mu$F. The frequency dependence of the reactive balance mentioned in Section 4.2 is illustrated in Figs. 15(*a*) and 15(*b*). Fig. 15(*a*) shows the variation in fundamental a.c. leak with frequency, when both P and C_2 are balanced for zero a.c. leak at one frequency only (1 kc/s). If this balance is carried out at 10 kc/s only, the curve of Fig. 15(*b*) results.

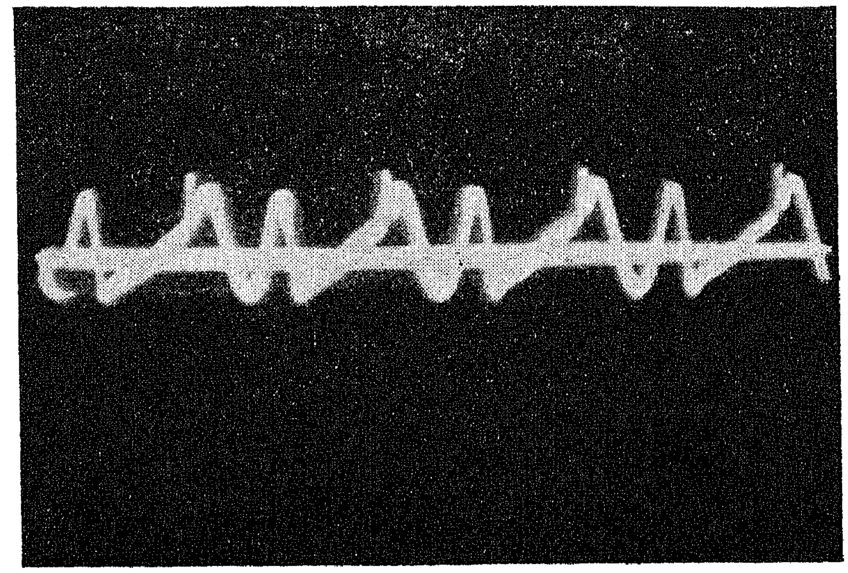
Zero leak can be obtained at all frequencies if both P and C_2 are balanced at each frequency, as in Fig. 15(*c*). The difference in the values of C_2 required to balance the modulator at 1 kc/s and 15 kc/s is approximately 50 $\mu\mu$F, i.e. of the order of 12·5%.

Fig. 15(*d*) shows the effect of balancing P and C_2 at 1 kc/s and rebalancing P alone for minimum leak at each subsequent frequency. This curve is very close to that for P and C_2 *both* balanced at 1 kc/s only [see Fig. 15(*a*)], apparently suggesting that misbalance of P is negligible in comparison with misbalance of C_2. This is not in fact the case, but it appears to be so since P is relatively insensitive when C_2 is badly misbalanced. As C_2 approaches balance the effect of P increases greatly. That both P and C_2 are of great importance is illustrated by Fig. 15(*e*), which shows the variation in leak when P is balanced at 1 kc/s alone, and C_2 rebalanced for minimum leak at each frequency.

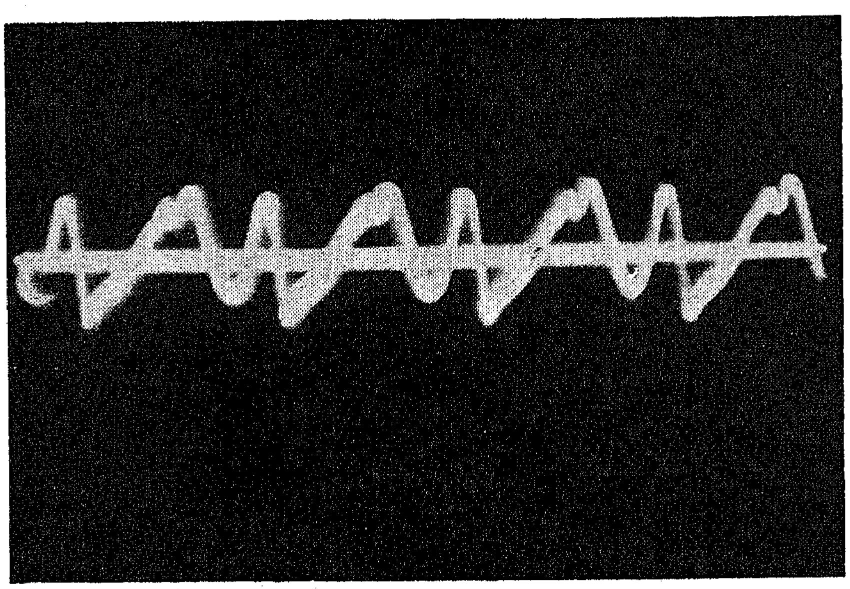
For Fig. 15(*f*), P and C_2 were balanced at 1 kc/s, then C_2 unbalanced by 50 $\mu\mu$F and P adjusted for minimum a.c. leak at each frequency. Note that the leak voltage is less than a millivolt over the range shown. The leak does not increase steadily as might at first be thought, owing to the capacitive misbalance, but rises at first and then falls to a low value as the frequency approaches that at which the new capacitance would give an a.c. balance.



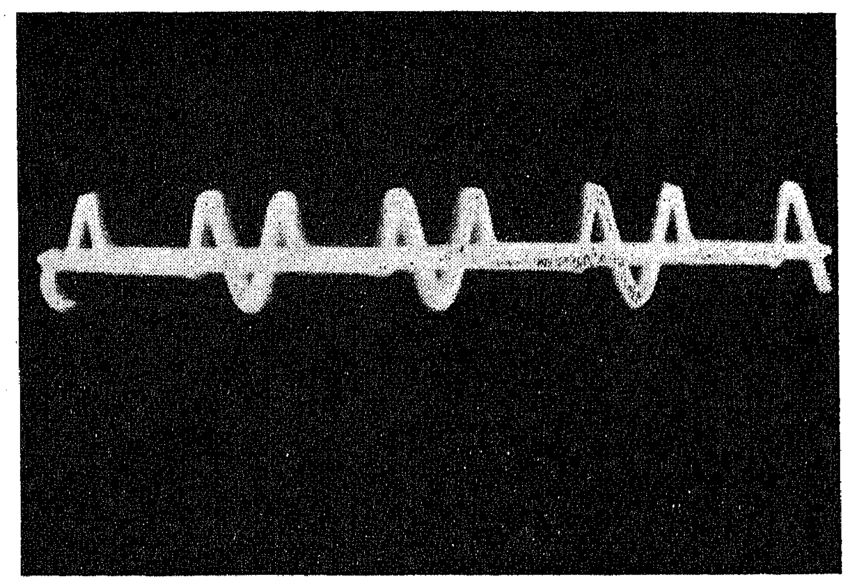
(a)



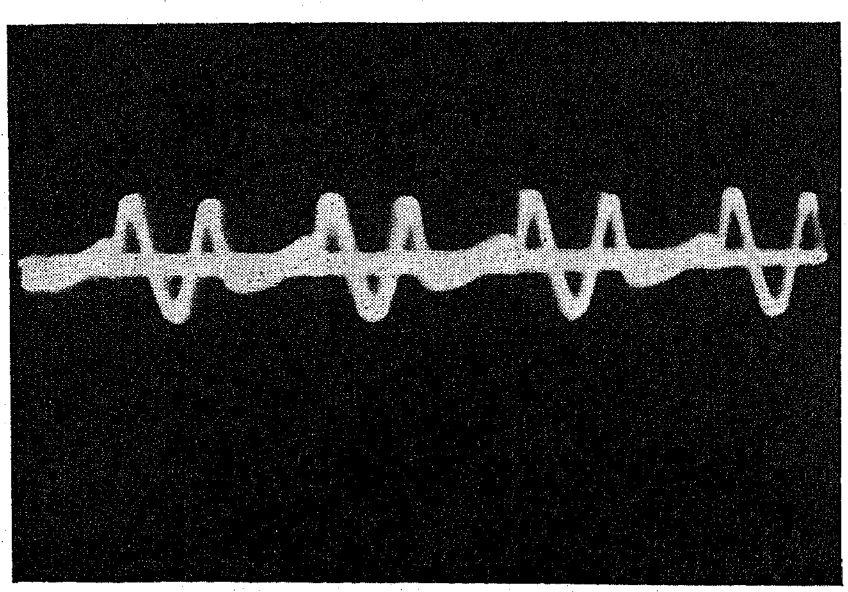
(d)



(b)



(e)

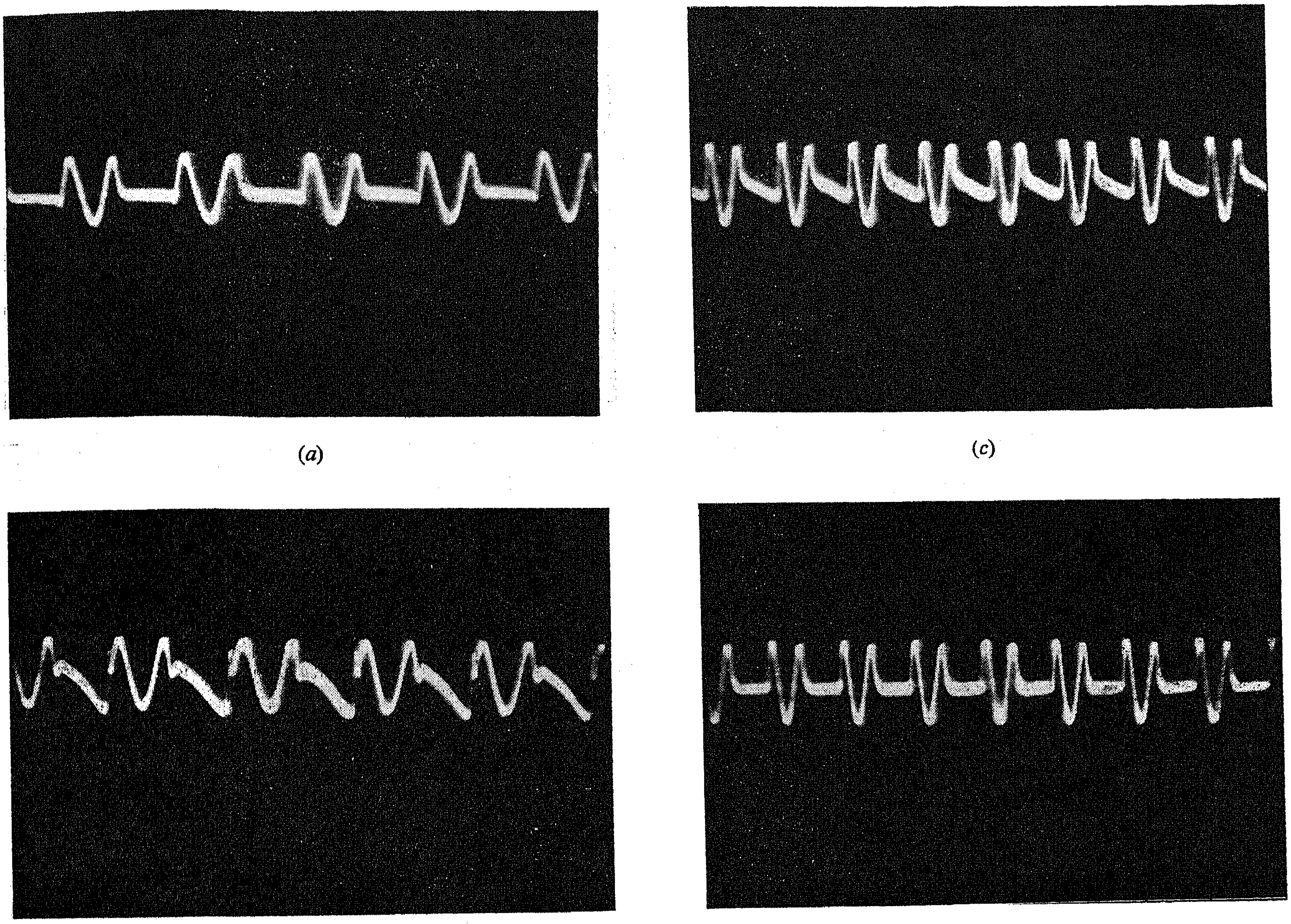


(c)

Fig. 13.—Effect of rectifier shunt capacitance on waveform of carrier-leak voltage.

Circuit as in Fig. 12.
$f = 1$ kc/s. Reference line at zero potential.

Waveform	Shunt capacitors connected	Fundamental a.c. leak
		mV
(a)	None	0·4
(b)	C_1	5·5
(c)	C_1 and C_3	1·6
(d)	C_1, C_2 and C_3	3·9
(e)	All	0·1



(a) (c)

(b) (d)

Fig. 14.—Effect of frequency and reactive balance on carrier-leak waveform.

Circuit as in Fig. 12.
Reference line at zero potential.

Waveform	Capacitors connected	Frequency	Fundamental
		kc/s	mV
(a)	All	1	0·05
(b)	C_1 removed	1	2·0
(c)	All	5	1·3
(d)	As (c) but with addition of 0–100 $\mu\mu$F variable capacitor across C_3 adjusted for minimum a.c. leak-voltage	5	0·05

(5.3.2) **Variation of D.C. Leak.**

The mean current corresponding to these curves is not constant, but in general falls with increasing frequency, reaching zero at some point around 10–15 kc/s, and thereafter going negative. Thus the correlation between a.c. and d.c. balance conditions varies, and in some cases the agreement is perfect at one frequency.

(5.3.3) **Modulator D.C.-Balanced.**

Fig. 16 shows similar curves for variation of a.c. leak with frequency, the modulator being balanced for zero d.c. leak. The procedure is similar to that for Fig. 15, except that P is adjusted for zero mean current instead of zero a.c. leak. Since it is always possible to do this, irrespective of the value of C_2, the latter was set at its a.c. balance value, i.e. in each case P and C_2 were balanced for zero a.c. leak, and P then readjusted to give zero mean current. With curve *e*, C_2 was left deliberately unbalanced by 50 $\mu\mu$F.

The curves of Fig. 16 will not be discussed in detail, but it will be noted that the leak over the frequency range taken is less than 3 mV. Also, owing to the effect mentioned in Section 5.3.2, the a.c. leak in some instances falls to a low value, corresponding to the fall in d.c. leak for the a.c.-balance condition.

When a reactive control is used, necessitating measurement of a.c. leak voltages, the d.c. balance technique might appear valueless. However, a process of initial accurate a.c. balancing followed by regular d.c. monitoring might well have application as suggested in Section 9.

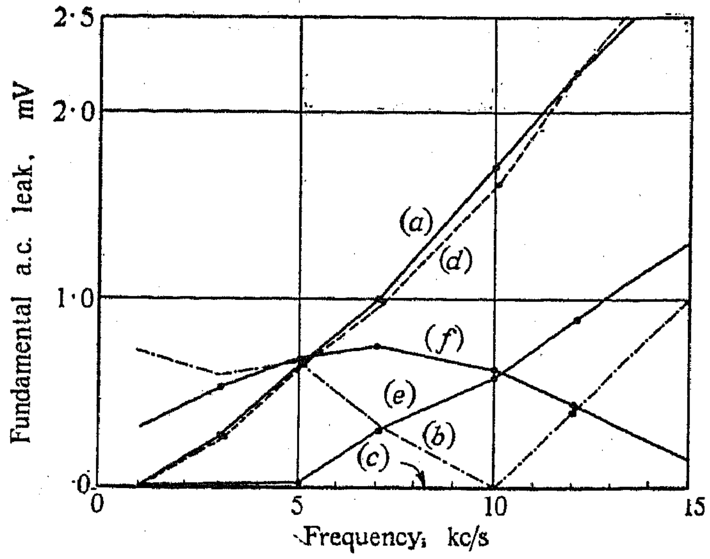


Fig. 15.—Variation of a.c. carrier-leak with frequency (modulator a.c.-balanced).

Circuit as in Fig. 12 but with C_2 variable.
(*a*) P and C_2 balanced at 1 kc/s only.
(*b*) P and C_2 balanced at 10 kc/s only.
(*c*) P and C_2 balanced at each frequency.
(*d*) P balanced at each frequency. C_2 balanced at 1 kc/s only.
(*e*) P balanced at 1 kc/s only. C_2 balanced at each frequency.
(*f*) P and C_2 balanced at 1 kc/s. C_2 then increased by 50 $\mu\mu$F. P balanced at each frequency.

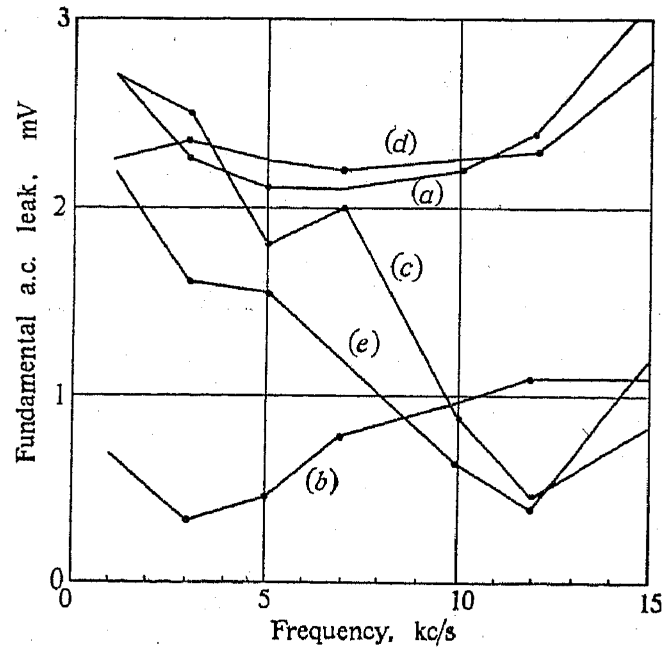


Fig. 16.—Variation of a.c. carrier leak with frequency (modulator d.c.-balanced).

Circuit as for Fig. 15.
(*a*) P and C_2 balanced for zero a.c. leak at 1 kc/s. P readjusted for zero d.c. leak at 1 kc/s only.
(*b*) As (*a*) at 10 kc/s.
(*c*) P and C_2 balanced for zero a.c. leak, and P readjusted for zero d.c. leak, at each frequency.
(*d*) P and C_2 balanced for zero a.c. leak at 1 kc/s, P only adjusted for zero d.c. leak at each frequency.
(*e*) P and C_2 balanced for zero a.c. leak at 1 kc/s, C_2 then increased by 50 $\mu\mu$F; P readjusted for zero d.c. leak at each frequency.

(5.4) Effect of Carrier-Generator Resistance

The effect of carrier-generator resistance was previously stated to be very small (Section 4.5 and Fig. 7), but it was noted that the minimum a.c. leak obtainable with a carrier-source resistance of 10 kilohms was higher than for lower values of source resistance. This is thought to be a reactive effect (the waveform is shown in Fig. 17; cf. Fig. 6) and will, of course, become greater at higher frequencies. Thus, to obtain a low carrier leak in the absence of a reactive balance, a low source resistance is desirable.

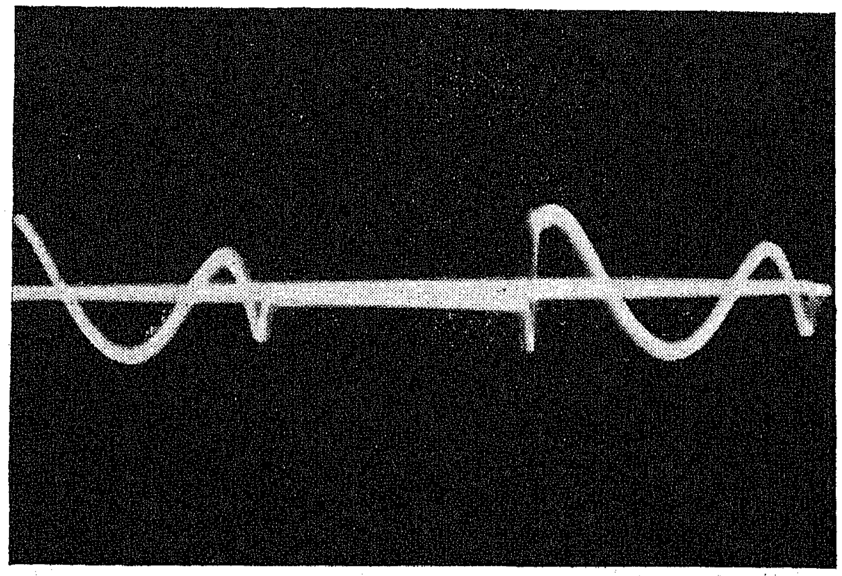


Fig. 17.—Effect of high carrier-generator resistance (R_C) on carrier-leak voltage waveform.

Circuit as in Fig. 3.
f = 3 kc/s, R_G = 10 kilohms, and R_L = 3·3 kilohms.
D.C.-balance condition. Fundamental a.c. leak = 2·2 mV.
Reference line at zero potential.

This is in contradiction to a statement in a previous paper[1] to the effect that the fundamental component of carrier-leak voltage decreases as the carrier-generator resistance is increased. This applies to the case of a non-reactive circuit and is therefore applicable only at low frequencies.

(5.5) Prediction of Carrier Leak at Other Frequencies

To study the variation of leak with capacitive unbalance, the circuit of Fig. 12 was used, with $C_2 = C_4 = 390\,\mu\mu$F (fixed), C_1 made up of eight 47 $\mu\mu$F capacitors in parallel, and C_3 made up of a 390 $\mu\mu$F capacitor in parallel with a 100 $\mu\mu$F trimmer. A known capacitive unbalance was obtained by removing a number of the 47 $\mu\mu$F capacitors comprising C_1. The relation between carrier-leak voltage and capacitive unbalance was found to be linear [Fig. 18(*a*)]. This is because the majority of the fundamental leak is due to capacitive unbalance. That the curve for a

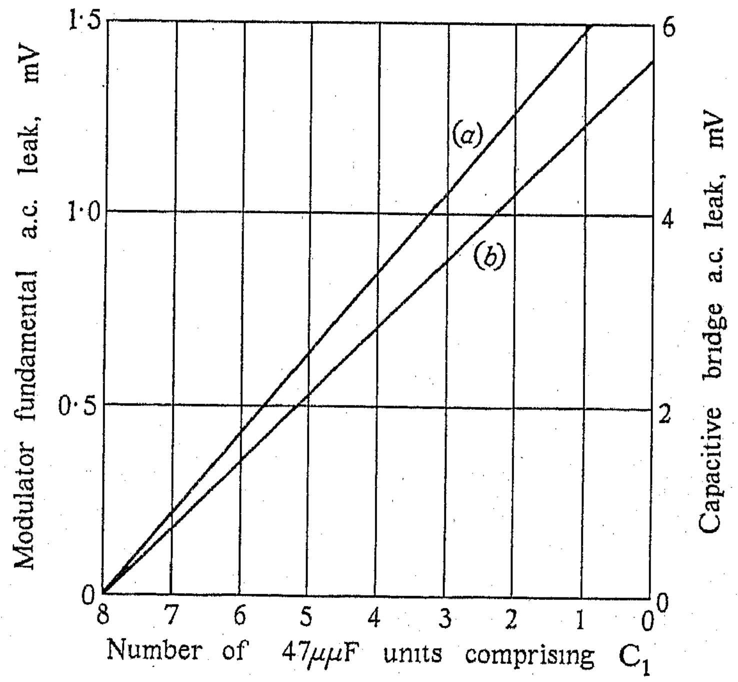


Fig. 18.—Variation of carrier-leak voltage with capacitive unbalance.

Circuit as in Fig. 12, with 0–100 $\mu\mu$F trimmer across C_3, and C_1 varied in units of 47 $\mu\mu$F.
C_3 and P balanced for zero a.c. leak with $C_1 = 8 \times 47\,\mu\mu$F.
f = 1 kc/s, R_G = 2·2 kilohms, and R_L = 3·3 kilohms. V_o = 1 volt forward peak.
(*a*) Modulator leak.
(*b*) Capacitive bridge leak (diodes rendered non-conducting by battery in series with carrier source).

purely capacitive and resistive bridge without time-varying elements is also linear is illustrated by Fig. 18(*b*), which was obtained by backing off the diodes with a large voltage in series with the carrier source, and applying 2·6 volts (r.m.s.) at 1 kc/s.

This suggests that, if the carrier-leak voltage for a given capacitive unbalance is known at one frequency, the leak at a higher frequency can be estimated. This was checked experimentally, and fair agreement obtained.

Fig. 19 shows carrier-leak voltage plotted against frequency for

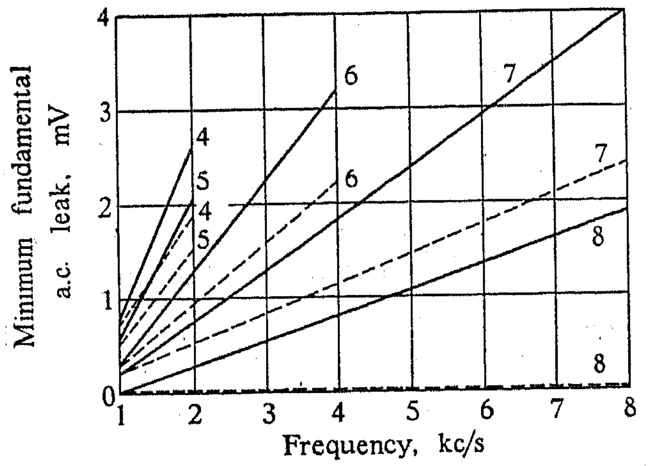


Fig. 19.—Linearity between minimum a.c. leak and frequency for a given capacitive unbalance.

Circuit as for Fig. 18.
Numbers on curves refer to number of 47 $\mu\mu$F units comprising C_1.
———— Trimmer adjusted at 1 kc/s only.
- - - - - - Trimmer readjusted at each frequency.

various degrees of capacitive unbalance. For the continuous curves the modulator was balanced for zero a.c. leak at 1 kc/s, and the trimmer left at that setting for subsequent frequencies. The broken curves were obtained by resetting the trimmer for zero a.c. leak at *each* frequency, prior to the removal of the 47 $\mu\mu$F units.

The two sets of curves can be compared by considering the conditions which should give the same carrier leak if a linear law is obeyed. For example, the leak voltage at 1 kc/s with four 47 $\mu\mu$F units removed should be the same as that at 2 kc/s with two removed, and that at 4 kc/s with one removed.

An example is set out in Table 1, from which it will be seen that the agreement is much improved when the trimmer is reset at each frequency.

Table 1

PREDICTION OF CARRIER LEAK AT OTHER FREQUENCIES

F	Number of 47 $\mu\mu$F units removed	A.C. leak	
		Trimmer set only at 1 kc/s	Trimmer reset at each frequency
kc/s		mV	mV
1	8	1·7	1·7
2	4	2·6	1·8
4	2	3·2	2·2
8	1	4·0	2·4
1	6	1·2	1·2
2	3	2·05	1·5
3	2	2·2	1·6
6	1	2·85	1·7

(6) TRANSFORMERLESS TYPE OF SHUNT MODULATOR[4,5]

This variation of the Cowan modulator may be used in circuits where the use of transformers is undesirable, often on account of phase-shift considerations. Measurements were made on the

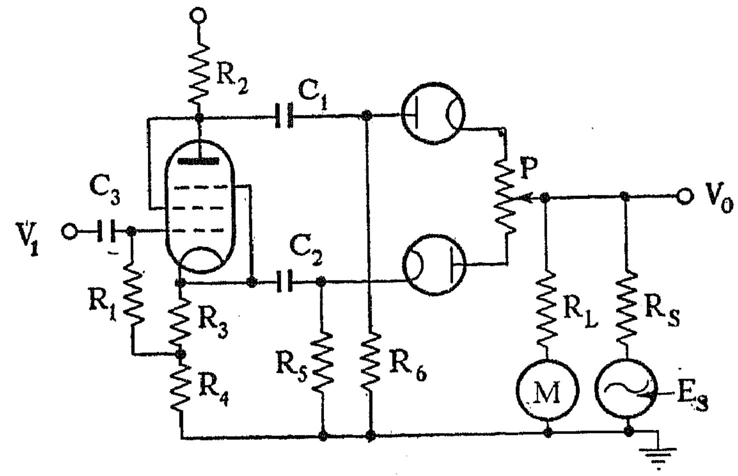


Fig. 20.—Transformerless type of shunt modulator.

Values used in experimental circuit:
C_3 = 0·01 μF, C_1 and C_2—various.
R_1 = 20 kilohms, R_2 = 2·2 kilohms, and R_3 = 200 ohms.
R_4 = 2·2 kilohms, R_5 = R_6 = 500 ohms.
P = 100 ohms. R_L = 3·3 kilohms.
Pentode: CV138. Diodes: CV140.
V_1 = Carrier voltage.
E_S = Signal voltage.
V_0 = Output voltage.

modulator of Fig. 20, using a carrier voltage of 1·5 volts (r.m.s.) at the phase-splitter input. This gives approximately 1 volt forward peak across each diode.

(6.1) Correlation between A.C. and D.C. Leak Voltages

As with the conventional shunt modulator previously described, correlation between a.c. and d.c. leak voltages was found to be sufficiently good for most practical purposes, i.e. where suppression of carrier voltage to less than 10 mV is satisfactory. Typical results gave a.c. leak voltages of the order of 5 mV or less for an input of 1·5 volts (r.m.s.).

(6.2) Variation with Voltage

Fig. 21 shows the variation of a.c. leak at 1 kc/s with carrier voltage, for a.c.-balanced and d.c.-balanced conditions (at

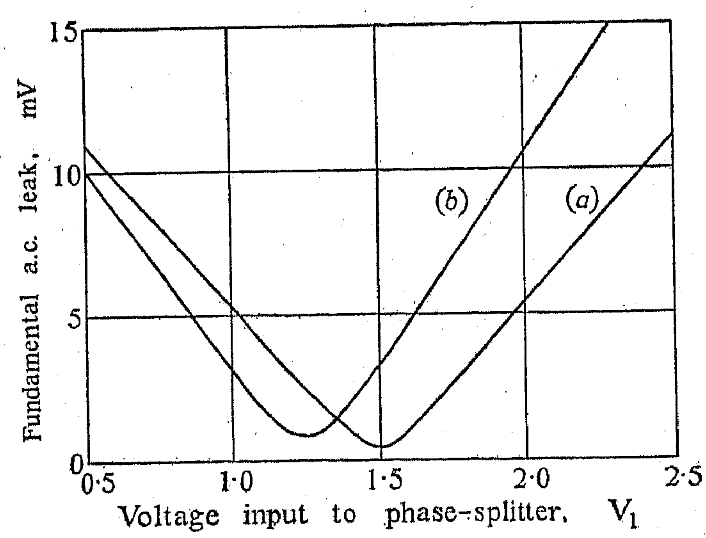


Fig. 21.—Variation of a.c. carrier leak of transformerless modulator with carrier voltage.

Circuit as in Fig. 20.
f = 1 kc/s, C_1 = C_2 = 16 μF.
(*a*) Modulator a.c.-balanced at V_1 = 1·5 volts.
(*b*) Modulator d.c.-balanced at V_1 = 1·5 volts.

1·5 volts). It will be noted that for an a.c. leak (fundamental) of less than 10 mV, an accuracy of carrier voltage of ±20% is sufficient.

(6.3) Effect of Coupling Capacitors

The reactance of the capacitors C_1 and C_2 in Fig. 20 must be kept low (of the order of a few ohms) at the frequency concerned

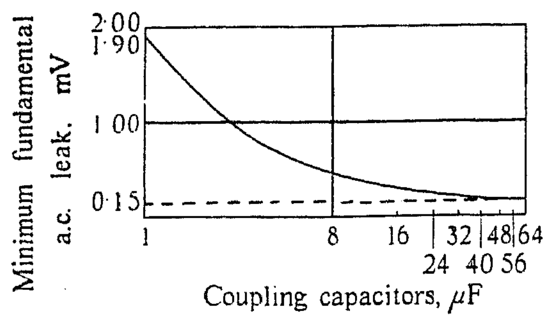


Fig. 22.—Variation of minimum a.c. leak with value of coupling capacitors.

Circuit as in Fig. 20.
$f = 1$ kc/s, $V_1 = 1 \cdot 5$ volts and $C_1 = C_2$.

if the a.c. leak is to be kept at a minimum. Fig. 22 shows the variation in leak with coupling capacitance for the modulator used. It will be seen that at the frequency employed (1 kc/s) coupling capacitors as large as, say, 16 μF (10 ohms) are required for low leak voltage.

A factor which must be taken into account if very low leak voltages are required is the effect of the differential time-constant of the two couplings. Fig. 23 shows the variation of minimum

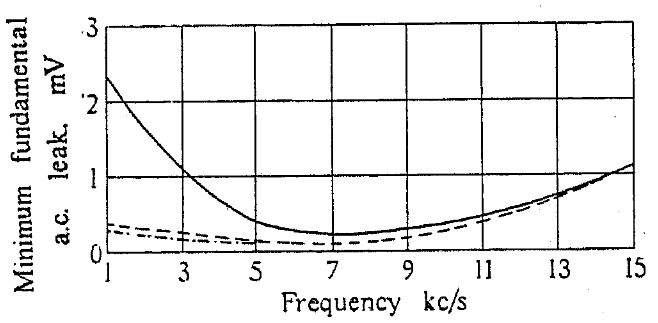


Fig. 23.—Variation of minimum a.c. leak with frequency, showing effect of value of coupling capacitors.

Circuit as in Fig. 20. $V_1 = 1 \cdot 5$ volts.
—— $C_1 = 8\,\mu$F, $C_2 = 16\,\mu$F.
---- $C_1 = C_2 = 8\,\mu$F.
—·— $C_1 = C_2 = 16\,\mu$F.

a.c. leak with value of coupling capacitance, and the effect of a large unbalance between C_1 and C_2. The leak rises at the low-frequency end of the curve owing to the effect of the reactance of the source (i.e. of C_1 and C_2) and at the higher-frequency end owing to reactive unbalance of the bridge network. Fig. 24 shows the effect on the waveform of the leak voltage.

Thus to a first approximation it would appear that the coupling capacitors should be equal and of low reactance.

Actually extremely low leak voltages (approaching zero) are

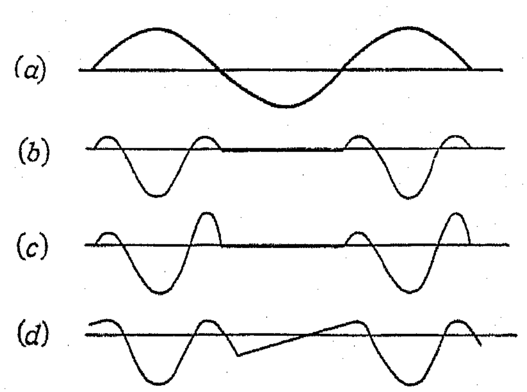


Fig. 24.—Waveforms of transformerless modulator carrier-leak voltage.

(a) Carrier voltage input: V_1.
(b) Carrier-leak voltage with $C_1 = C_2$ and no large reactive unbalance of rectifier-resistance bridge.
(c) Effect of 2 : 1 inequality of coupling capacitors.
(d) Effect of large reactive unbalance of bridge.
(Compare Figs. 13 and 14.)

obtained by making C_1 greater than C_2. This effect has not yet been investigated in detail, but is most probably due to the difference in source impedance at the anode (approximately R_2) and at the cathode (approximately $1/g_m$). In one experiment the leak voltage at 1 kc/s with $C_1 = C_2 = 7 \cdot 7\,\mu$F (measured) was reduced from about 300 μV to less than 5 μV by the addition of 1 μF across C_1.

Thus a preferable alternative to the use of the large capacitors (16 μF) mentioned previously is the use of much smaller values, with provision for varying one to achieve minimum leak.

(7) THE USE OF GERMANIUM-CRYSTAL RECTIFIERS IN MODULATORS

Germanium diodes are now manufactured covering a wide range of characteristics. Rectifiers in a typical commercial range, for example, are as follows:

Forward resistance at 1 volt forward peak	1 000, 500 or 200 ohms.
Backward resistance at −10 volts..	10, 100, 300 or 1 000 kilohms.
Peak working reverse voltage ..	20, 40, 60, 80 or 100 volts.

Germanium rectifiers are nowadays reasonably rugged and reliable, and have the added advantages of compactness and low self-capacitance (approximately 1 $\mu\mu$F), and, of course, they require no heater supplies.

The performance of a conventional shunt modulator as previously described, but using Westinghouse WG7A germanium rectifiers, was found to be very similar to that using thermiomic diodes CV140.

(8) STABILITY OF CARRIER LEAK

(8.1) Factors affecting the Maintenance of Low Carrier-Leak Voltages

It has been shown that carrier-leak voltages of a very low order (with care, almost zero) can be realized by the use of reactive as well as resistive balancing of the modulator bridge network. The maintenance of the leak at these low levels, however, is an extremely difficult matter. Variables which have to be accurately controlled include:

(a) Temperature (variation of which varies rectifier characteristics).
(b) Heater current (in the case of thermionic diodes).
(c) Carrier voltage.
(d) Carrier frequency.

Also, since rectifier characteristics vary with age, frequent monitoring is essential.

(8.2) Constant-Current Ring Modulator

Stability of carrier leak in a ring modulator can be considerably improved by using a circuit suggested by Cooper[8] and investigated experimentally by Tucker,[5] but no analogous method seems possible with the series or shunt modulators. The basis of the method is the independent control of the currents in the two halves of the output transformer primary winding, unbalance of which gives rise to carrier leak (as opposed to the *voltage* unbalance of the shunt-type modulators). This is achieved by the use of separate constant-current generators for the carrier supply.

If a square-wave carrier is used, very low leak levels can be realized and maintained. With a sinusoidal carrier, unbalance occurs during the transition period, when the constant currents divide, not necessarily equally, between the series and shunt rectifiers of the lattice. However, leak voltages of the same order as those of the basic circuits can be obtained, with the added advantage of greatly improved stability.

Since the parameter of primary importance for most modulator applications is dynamic range (i.e. the ratio of maximum signal

output to carrier leak), it should be remembered that the ring modulator has the additional advantage of greater conversion efficiency.

(9) DISCUSSION OF RESULTS AND CONCLUSIONS

(9.1) Square-Wave Carrier

The use of a square-wave carrier e.m.f. in a low-frequency—i.e. non-reactive—circuit gives a good approximation to ideal switching conditions (i.e. where the rectifiers change from a constant forward resistance to a constant backward resistance, and vice versa, instantaneously and at zero voltage). Carrier-leak voltage is then due primarily to unbalance of rectifier forward resistances and is quite calculable. Also, good suppression of the total carrier-leak voltage (as opposed to the fundamental component only) can be obtained using a balancing resistor. Experimental results obtained agreed closely with calculated values, and d.c./a.c. correlation is excellent if rectifiers with high or alternatively well-balanced back resistances are used.

(9.2) Sinusoidal Carrier

(*a*) With a sinusoidal carrier e.m.f. the concept of ideal switching can no longer be applied, and the calculation of carrier leak becomes too involved to be practicable. Leak voltages arise mainly owing to unbalance of the rectifier bridge during the transition period between forward- and backward-resistance conditions. The total leak voltage cannot therefore be controlled so effectively by means of a balancing resistor as in the square-wave case. In a perfectly non-reactive circuit, of course, complete elimination of any one component is theoretically possible by means of a resistive balance.

Measured results at low frequencies (less than 5 kc/s) were inferior compared with those obtained using a square-wave carrier. The total leak was much greater owing to the effect described above, and stray reactance, although small, prevented the complete suppression of the fundamental component. In the experimental circuit it was difficult to obtain, by resistance balancing, fundamental components of the leak voltage lower than $0 \cdot 1\%$ of the carrier forward peak potential difference.

The d.c./a.c. correlation (in spite of high back-resistances) was such that zero d.c. leak might correspond to anything up to $0 \cdot 4\%$ leak of fundamental a.c. component.

(*b*) By employing a reactive as well as a resistive balance in these low-frequency circuits, however, it was possible to reduce the fundamental leak to zero order (i.e. less than $5\,\mu$V) for short periods. No work has as yet been done on the maintenance of these very low leak voltages, but it is thought that even on a long-term basis, the leak voltage should in many instances be considerably lower (say 10–20 dB) than that without a reactive balance.

(*c*) At higher frequencies it has been demonstrated that unbalance of rectifier and stray reactances contributes very largely to the carrier leak, but that results as good as those of (*a*) can be obtained by balancing the reactance of the circuit.

Reactive and resistive balance should therefore be very useful where leak voltages are required to be lower than normally obtained.

Correlation between a.c. and d.c. leaks is quite reasonable with a reactive unbalance. Even with an unbalance as large as $50\,\mu\mu$F at 1 kc/s (equivalent to the quite likely value of $0 \cdot 5\,\mu\mu$F at 100 kc/s), the d.c./a.c. correlation was such that for zero d.c. leak the a.c. leak did not exceed about $0 \cdot 5\%$ of the carrier forward peak potential difference.

(*d*) Under these reactive conditions, the effect of carrier-generator impedance is complicated, and even if it is resistive, the higher values may give greater leak (not less, as with non-reactive or low-frequency circuits) than the lower values. If the carrier-generator impedance is reactive, the capacitance balance becomes frequency dependent, i.e. a different adjustment of the balancing capacitor is needed at each frequency. This dependence on the carrier-generator impedance probably exists because the rectifier bridge is not truly balanced over the whole switching cycle, but is merely adjusted to give minimum output of one particular component.

(9.3) Use of D.C. Balance

It seems a fairly safe general conclusion that control of carrier-leak adjustment by the use of a d.c. meter is likely to prove very useful in practice, and provided that efficient rectifiers are used, it should enable fundamental a.c. leak voltages to be maintained at least 40 dB below the carrier voltage across the rectifiers. The main attraction of this method, of course, lies in the simplicity of regular monitoring of carrier leak.

With reactive circuits, or at higher frequencies, a lower leak voltage than normal might be maintained by an initial accurate adjustment of a.c. leak (involving reactive balance and a.c. measurement) and subsequent d.c. monitoring, since ageing will affect the rectifier resistances, although not to any large extent their reactances. This scheme has not as yet been tried in practice.

(10) ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Dr. D. G. Tucker and Mr. J. W. R. Griffiths for advice and encouragement during the experimental work and the preparation of the paper, and to Dr. J. H. Mole for his very helpful criticism and advice.

The paper is published by permission of the Admiralty, and expresses solely the personal opinions of the author.

(11) REFERENCES

(1) TUCKER, D. G.: "Some Aspects of the Design of Balanced Rectifier Modulators for Precision Applications," *Journal I.E.E.*, 1948, **95**, Part III, p. 161.

(2) TUCKER, D. G.: "Carrier-Leak Control in Rectifier Modulators by the Use of a D.C. Meter," *Proceedings I.E.E.*, Paper No. 1384 R (Part 2), November, 1952 (**99**, Part III, p. 402).

(3) TUCKER, D. G.: "Rectifier Resistance Laws," *Wireless Engineer*, 1948, **25**, p. 117.

(4) TUCKER, D. G.: "Balanced Rectifier Modulators without Transformers," *Electronic Engineering*, 1950, **22**, p. 139.

(5) TUCKER, D. G.: "Modulators and Frequency-Changers" (Macdonald, London, 1953). Includes all the above except Reference 2.

(6) MOULLIN, E. B.: "The Problem of a Non-Ohmic Resistor in Series with an Impedance," *Proceedings I.E.E.*, Paper No. 1048, March, 1951 (**98**, Part I, p. 87).

(7) SLEMON, G. R.: "A Method of Approximate Steady-State Analysis for Non-Linear Networks," *ibid.*, Paper No. 1530, September, 1953 (**100**, Part I, p. 275).

(8) COOPER, W. H. B.: Discussion on paper by D. G. Tucker.[1] *Ibid.*, 1949, **96**, Part III, p. 215.

(9) TUCKER, D. G.: "Unbalance Effects in Modulators," *Journal of the British Institution of Radio Engineers*, 1955, **15**, p. 199.

621.372.8

Paper No. 1852 R
July 1955

A DOUBLE-GROUND-PLANE STRIP-LINE SYSTEM FOR MICROWAVES

By BENGT A. DAHLMAN.

(*The paper was first received* 3*rd December*, 1954, *and in revised form* 21*st February*, 1955.)

SUMMARY

The double-ground-plane strip line consists of two parallel conducting planes with a conducting strip imbedded in a homogenous dielectric between them. Transmission characteristics for this system are calculated, and design formulae are given. Practical viewpoints on the design and application of strip lines are discussed. The system can be used as an inexpensive base for microwave circuits and is well adapted both to laboratory experiments and mass production.

LIST OF SYMBOLS

ϵ_0 = Absolute permittivity of free space, farads/m.
ϵ = Relative permittivity of dielectric in the line.
δ = Dielectric loss angle.
σ = Conductivity, mhos/m.
μ_0 = Absolute permeability of free space, henrys/m.
ε = Base of natural logarithms.
λ_0 = Wavelength for transmission in free space, m.
λ = Wavelength along the line, m.
f = Frequency, c/s.
Z_0 = Characteristic impedance, ohms.
C = Capacitance per unit length of the line, farads/m.
b = Width of strip, m.
$2h$ = Distance between ground planes, m.
d = Thickness of strip, m.
t = Width of ground planes, m.
E = Electric force, volts/m.
E_h = Homogeneous electric force far from the edge of a very wide strip, volts/m.
P_c = Power loss per unit length of the conductors, watts/m.
P_T = Power transmitted along the line, watts.
α_d = Dielectric attenuation, dB/wavelength.
α_c = Conductor attenuation, dB/m.
V = Potential difference between the ground planes and the strip, volts.

(1) INTRODUCTION

During the last few years there has been considerable interest in new simpler methods for manufacturing microwave circuits. In December, 1952, the Federal Telecommunication Laboratories presented an extensive report on a microwave printed-line system.[1–3] Although it is simple enough and has proved to be very useful for many circuits, it has the disadvantage of being an open system and is thus subject to some radiation. A shielded strip-line system was described by Barrett and Barnes,[4] but no real analysis was given.

The double-ground-plane strip line described in the paper has a thin strip of copper foil placed between two sheets of low-loss dielectric, and the outer sides of the dielectric sheets are covered with conducting planes.

The following analysis provides a theory for the double-ground-plane strip line similar to that given in Reference 2.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Dahlman is with Svenska A.B. Trådlös Telegrafi, Sweden.

(2) TRANSMISSION CHARACTERISTICS OF THE DOUBLE-GROUND-PLANE STRIP LINE

Interest is mainly centred on the phase wavelength, characteristic impedance and attenuation of the strip line.

(2.1) Phase Wavelength

A cross-section of the line is shown in Fig. 1. Since the electric field does not extend outside the dielectric, the fundamental will be a pure TEM mode. The phase wavelength in the strip line is then $1/\sqrt{\epsilon}$ times the free-space wavelength, where ϵ is the dielectric constant of the material in the line:

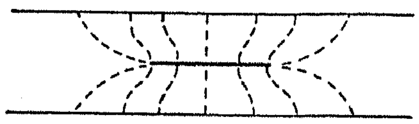


Fig. 1.—Cross-section of the double-ground-plane strip line.

$$\lambda = \lambda_0/\sqrt{\epsilon} \qquad (1)$$

(2.2) Characteristic Impedance

For a TEM mode the characteristic impedance, Z_0, is given by

$$Z_0 = \sqrt{(\mu\epsilon_0\epsilon)}/C \qquad (2)$$

where C is the capacitance per unit length. This capacitance is most easily calculated by the conformal mapping method.

An infinitely thin strip is assumed. As the subsequent analysis will show, the electric strength between the strip and the conducting planes is very nearly equal to the homogeneous electric strength between two parallel conducting planes of infinite width with the same potential difference and the same spacing, except very close to the strip edges. We can therefore, to a good approximation, calculate the field at the edge of the strip by assuming that the strip extends to infinity in one direction.

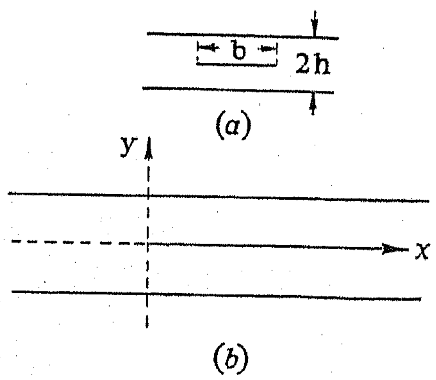


Fig. 2.—The actual strip and the approximation.

Figs. 2(*a*) and 2(*b*) show the actual strip and the approximation. The co-ordinate axes of a rectangular co-ordinate system in the z-plane are shown in Fig. 2(*b*).

By means of the transformation

$$z = \frac{2h}{\pi}\log z_1 \qquad z_1 = x_1 + jy_1 \qquad (3)$$

the strip is transformed to the x_1-axis from $x_1 = 1$ to $x_1 = \infty$ in the z_1-plane and the ground planes to the y_1-axis.

The z_1-plane is shown in Fig. 3. The field in the z_1-plane is given in several textbooks, the lines of force being ellipses, but to get a better idea of the problem we shall make one more transformation. This will prove to be useful later in the calculation of the attenuation.

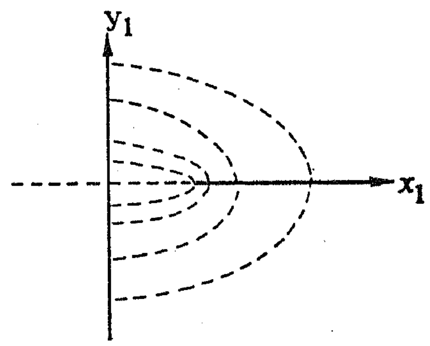


Fig. 3.—The double-ground-plane strip line transformed to the z_1-plane.

The transformation

$$z_1 = \cos \omega \qquad \omega = u + jv \quad . \quad . \quad . \quad . \quad (4)$$

transforms the strip to the v-axis and the ground planes to the line $u = \pi/2$ in the ω-plane.

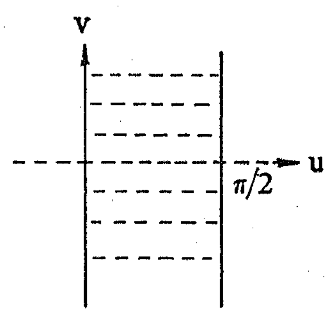


Fig. 4.—The double-ground-plane strip line transformed to the ω-plane.

The centre line of the actual strip corresponds to the point $(b/2, \theta)$ in the z-plane and to

$$\omega_p = u_p + jv_p$$

in the ω-plane. ω_p is given by eqns. (3) and (4).

$$\left.\begin{aligned} u_p &= 0 \\ v_p &= \pm \text{ arc cosh } \varepsilon^{\pi b/4h} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The capacitance per unit length between the v-axis and the line $u = \pi/2$ from $v = -|v_p|$ to $v = |v_p|$ is

$$C = \epsilon\epsilon_0 \frac{\text{arc cosh } \varepsilon^{\pi b/4h}}{\pi} \times 4 \quad . \quad . \quad . \quad . \quad (6)$$

As this represents the capacitance of half of the strip, the impedance given by eqn. (2) is

$$Z_0 = \sqrt{(\mu_0\epsilon\epsilon_0)}/8 \times \epsilon\epsilon_0 \frac{\text{arc cosh } \varepsilon^{\pi b/4h}}{\pi} \quad . \quad . \quad (7)$$

In Fig. 5, Z_0 is plotted against h/b. An approximate expression for the impedance of a strip of finite thickness d can easily be obtained by assuming that the image of the strip in the ω-plane is the line $u = \Delta u$ (see Fig. 6). The image of this contour in the z-plane is shown in Fig. 7. The assumption will be made that $d \ll h$. The equation for the strip contour in the z-plane given by eqns. (3) and (4) is then

$$y^2 = \left(\frac{d}{2}\right)^2 \times (1 - \varepsilon^{-x\pi/h}) \quad . \quad . \quad . \quad . \quad (8)$$

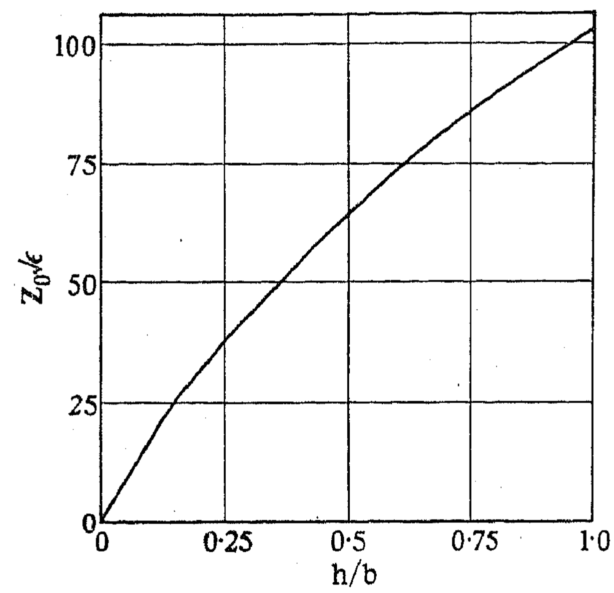


Fig. 5.—Characteristic impedance of double-ground-plane strip line.

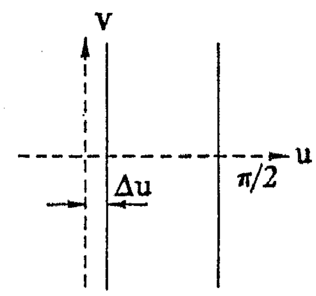


Fig. 6.—Image of strip of finite thickness.

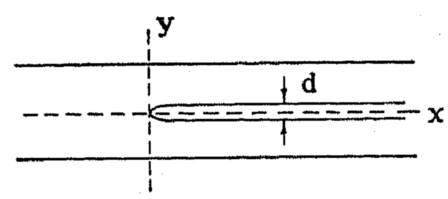


Fig. 7.—Strip of finite thickness.

and the value of Δu is

$$\Delta u = \frac{\pi}{4} \times \frac{d}{h} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

For thin strips the approximation for the actual strip cross-section obtained in this way is probably as good as any. To get the impedance of a strip of finite thickness we should multiply the value of the impedance obtained from eqn. (7) and Fig. 5 by a factor

$$Dz = 1 - \frac{d}{2h} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The approximation entailed in assuming that the strip extends to infinity in one direction will next be checked. If the electric field in the ω-plane is $E(\omega)$ the field in the z-plane is given by

$$E(z) = E(\omega) \times \overline{d\omega/dz} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where the bar indicates the complex conjugate.

Thus, using eqns. (3) and (4),

$$E(z) = E(\omega)\frac{d}{dz}(\text{arc cos } \varepsilon^{\pi z/2h}) = E(\omega)\overline{\frac{\pi}{2h\sqrt{(\varepsilon^{-zh/\pi} - 1)}}} \quad . \quad (12)$$

The ratio between the actual field strength at the surface of the strip and the homogeneous field strength far from the edge is plotted in Fig. 8. It is thus obvious that for $b/h > 1 \cdot 5$ the approximation is accurate to within a few per cent. Even for a

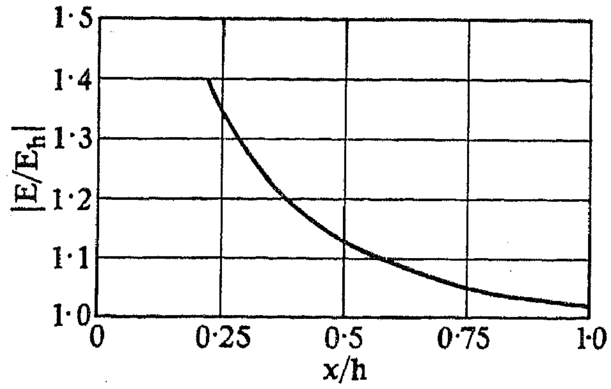


Fig. 8.—Variation of the electric field across the strip.

ratio of $b : h$ as low as unity, the value of Z_0 given in eqn. (7) and Fig. 5 is only 3% bigger than the exact value calculated in Reference 11.

(2.3) Attenuation

The attenuation is due to dielectric losses and losses in the conductors. For TEM-mode propagation the attenuation due to dielectric losses is given by the following expression:

$$\alpha_d = 20 \times \frac{\delta\pi}{\log 10} = 27 \cdot 2\delta \text{ decibels per wavelength} \quad . \quad (13)$$

The conductor losses per unit length of the line are

$$P_c = k\int |E(z)|^2 |dz| = k\int |E(\omega)|^2 \left|\frac{d\omega}{dz}\right| |d\omega| \quad . \quad . \quad (14)$$

where

$$k = \sqrt{\left(\frac{\pi f}{\sigma\mu_0}\right)} \times \frac{\epsilon\epsilon_0}{2} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The integral should be taken along the conductor contours in the z- or ω-plane. The approximation for an actual strip of thickness d is used as described in the previous Section and shown in Fig. 8. The integration is made in Appendix 7.1.

$$P_c = \frac{4V^2k}{h}\left(\frac{b}{h} + \frac{2}{\pi}\log\frac{4h}{\pi d}\right) \quad . \quad . \quad . \quad . \quad (16)$$

The power transmitted along the line is

$$P_T = \frac{V^2}{2Z_0}$$

The attenuation per unit length due to conductor losses is

$$\alpha_c = \frac{10}{\log 10} \times \frac{P_c}{P_T} = \frac{80Z_0k}{2 \cdot 3h}\left(\frac{b}{h} + \frac{2}{\pi}\log\frac{4h}{\pi d}\right) \quad . \quad (17)$$

The value of k obtained from eqn. (15) is substituted in eqn. (17) and we obtain

$$\alpha_c\sqrt{\frac{\sigma}{f\epsilon\epsilon_0}}\,h = 0 \cdot 082Z_0\sqrt{\epsilon}\left(\frac{b}{h} + \frac{2}{\pi}\log\frac{4h}{\pi d}\right) \quad . \quad (18)$$

Eqn. (18) is illustrated in Fig. 9.

A 50-ohm strip line with polystyrene dielectric at a frequency of 3 000 Mc/s is considered:

$$\alpha_d = 27 \cdot 2 \times 0 \cdot 001 = 0 \cdot 027 \text{ dB/wavelength}$$
$$\lambda = 10/\sqrt{2 \cdot 5} \simeq 6 \cdot 3 \text{ cm}$$

If the conductor losses are to be smaller than the dielectric losses,

$$\alpha_c < 0 \cdot 027/0 \cdot 063 = 0 \cdot 43 \text{ dB/m}$$

$$\sqrt{\frac{\sigma}{f\epsilon\epsilon_0}} = 29\,000$$

$$h/b = 0 \cdot 65,\ b/h = 1 \cdot 55$$

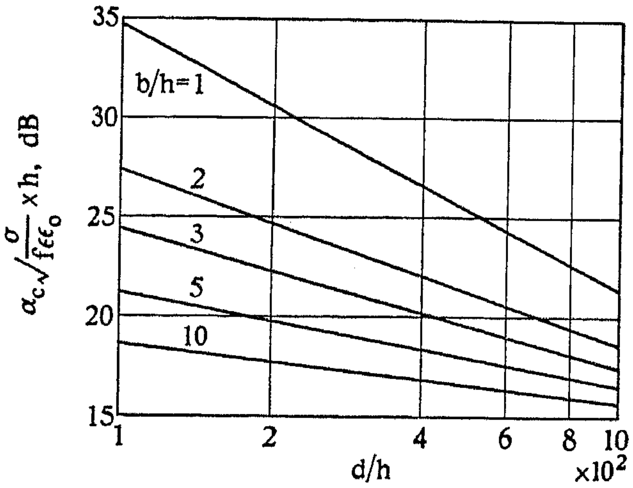


Fig. 9.—Attenuation due to conductor losses.

If we choose

$$h = 2\text{ mm}$$
$$\alpha_c\sqrt{\frac{\sigma}{f\epsilon\epsilon_0}} \times h = 25 \text{ dB/m}$$
$$d/h = 0 \cdot 03$$
$$d = 0 \cdot 06 \text{ mm}$$

(3) PRACTICAL VIEWPOINTS ON THE DESIGN AND APPLICATION OF STRIP-LINE SYSTEMS

(3.1) Dimensions

Various factors such as impedance, the maximum permissible attenuation, power-handling capacity and available space determine the practical dimensions of strip-line systems. In the previous Section, impedance and attenuation characteristics were considered. The power-handling capacity is determined by the power dissipation allowed in the line, and by the electric strength of the insulation.

The power dissipation per unit length of the line is

$$P_l = P_T\alpha\frac{\log 10}{10} = 0 \cdot 23 \times P_T\alpha \text{ watts per unit length} \quad . \quad (19)$$

where P_T is the transmitted power and α is the total attenuation in decibels per unit length [see eqns. (13) and (18)].

The strongest electric field occurs at the strip edges. The position of the strip edge in the z-plane (see Fig. 8) is

$$\left[\frac{-\pi}{h}\left(\frac{d}{4}\right)^2,\ 0\right]$$

which corresponds to the point $(\Delta u, 0)$ in the ω-plane. From eqn. (12) the electric field at the edge is then

$$E_{max} = \frac{4}{\pi}\frac{V}{d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Even though the calculations made so far are based on the assumption of ground planes with infinite extension, the results are valid with good approximation even for finite ground planes, so long as the electric field at the edge of the planes is small compared with the homogeneous field between the ground planes and the strip. The variation of the electric field across the ground planes, given by eqn. (12), is shown in Fig. 10.

A ground-plane width of

$$t = b + 4h$$

is adequate for most purposes.

(3.2) Manufacturing Methods

Double-ground-plane strip-line systems can be manufactured by methods similar to those used in the technique described in

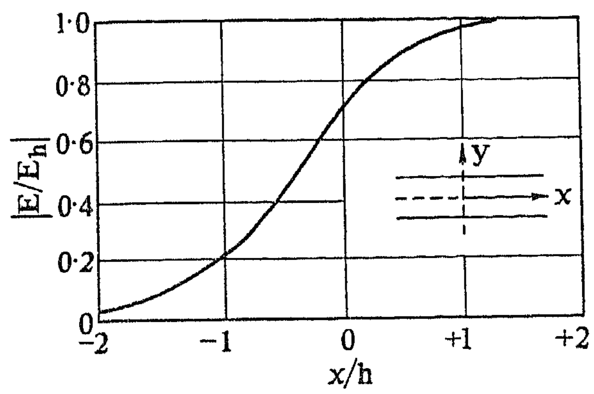


Fig. 10.—Variation of the electric field across the ground planes.

References 1 and 5. For mass production the best method so far found is a process by which the strip configuration is printed on one side of a copperclad dielectric sheet. The excess material is removed by an etching process. An unclad dielectric sheet of the same thickness is then laid on top of the strip and the sandwich is placed between two suitable ground-plane conductors. If one of the dielectric sheets is clad on both sides and the other on one side the copper skins may serve as ground planes.

For laboratory use a convenient way of making strip-line components is to cut out the desired strip configuration in copper foil and place it between two dielectric sheets.

(3.3) Suitable Dielectrics

Low-loss copperclad dielectrics are commercially available. For laboratory purposes unclad polystyrene and Teflon sheets are often practicable.

(3.4) Advantages of Strip-Line Systems

Microwave circuits in the conventional waveguide and coaxial systems are often mechanically complicated and rather expensive. This is especially true of networks with more than four terminals —the costs rise very rapidly with the number of terminals. For such types of circuits the strip-line technique can be used to reduce the costs considerably. In a practical example, Fig. 11 shows a

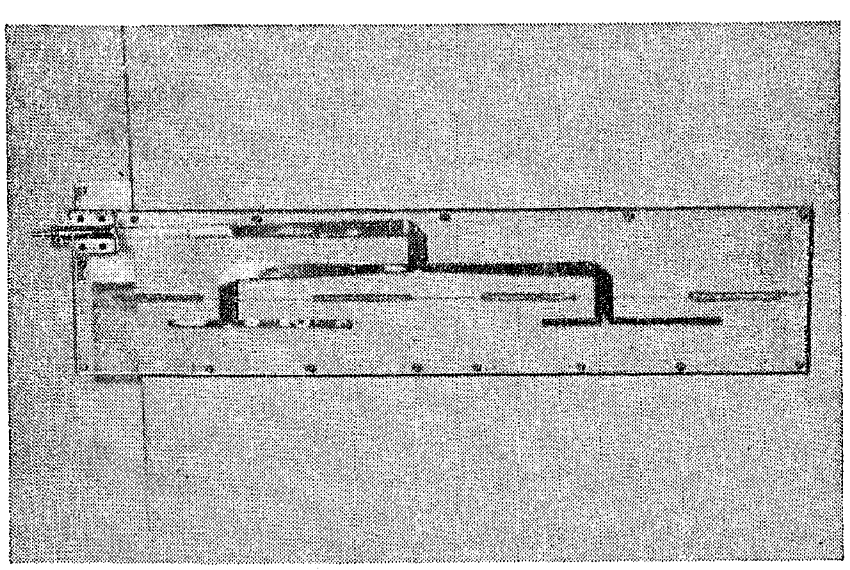


Fig. 11.—An antenna feeder system in a strip line.

feeder system for a certain antenna. The whole system involving 50/25-ohm tapers, right-angle bends, tee junctions and an adaptor to a type-N female connector is matched throughout at about 3 000 Mc/s. The dielectric is 2 mm thick polystyrene, which gives an attenuation of about 0·03 dB per wavelength. The strip is 0·04 mm thick copper foil, which gives a conductor attenuation of about 0·5 dB/m. The whole system has an attenuation of about 0·3 dB. If the circuit were to be constructed in a coaxial system the circuit configuration would probably be cut out in two solid metal blocks with a milling machine. The two blocks put together would make the outer conductor. The inner conductor would have to be supported by some dielectric supports which should be compensated. The long tapers which are easily made in the strip system would probably have to be substituted by transformers. An attempt has been made to estimate the production costs for the two systems. The comparison is given in Table 1.

Table 1

Number of units	Price per unit	
	Coaxial line	Strip line
1	£15	£2
25	£5	15s.
1 000	10s.	2s.

The costs for the strip-line units should be 10–20% of those of the coaxial units.

(4) CONCLUSIONS

The double-ground-plane strip line is a useful complement to the conventional waveguide and coaxial systems and to other types of strip lines. Complicated networks can be built in this system at much lower cost than in the more conventional waveguide and coaxial systems and they do not have the disadvantages of unshielded systems.

(5) ACKNOWLEDGMENTS

The author wishes to express his thanks to the Research Institute of National Defence, which supported the project.

(6) REFERENCES

(1) GRIEG, D. D., and ENGELMANN, H. F.: "Microstrip—A New Transmission Technique for the Kilomegacycle Range," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1644.
(2) ASSADOURIAN, F., and RIMAI, E.: "Simplified Theory of Microstrip Transmission Systems," *ibid.*, p. 1651.
(3) KOSTRIZA, J. A.: "Microstrip Components," *ibid.*, p. 1658.
(4) BARRETT, R. M., and BARNES, M. H.: "Microwave Printed Circuits," *Radio and TV News, Radio Electronic Engineering Section*, 1951, **46**, pp. 16 and 31.
(5) BARRETT, R. M.: "Etched Sheets save as Microwave Components," *Electronics*, June, 1952, p. 114.
(6) FUBINI, E. G., FROMM, W., and KEEN, H.: "New Techniques for High-Q Microwave Components," *Convention Record of the Institute of Radio Engineers, 1954 National Convention*, Part 8, p. 91.
(7) COHN, S.: "Characteristic Impedance of the Shielded-Strip Transmission Line," *Transactions of the Institute of Radio Engineers*, July, 1954, **MTT-2**, p. 52.
(8) COHN, S.: "Problems in Strip Transmission Lines," presented at the Tufts College Symposium on Microwave Strip Circuits on 12th October, 1954 (to be printed in forthcoming Transactions of the PG-MTT).
(9) PACKARD, K. S.: "Machine Methods make Strip Transmission Line," *Electronics*, September, 1954, p. 149.

(10) OLIVER, A. A.: "The Radiation Conductance of a Series Slot in Strip Transmission Line," *Convention Record of the Institute of Radio Engineers, Communications and Microwaves*, 1954, Part 8, p. 89.

(11) OBERLETTINGER, F., and MAGNUS, W.: "Anwendung der Elliptischen Funktionen Physik und Technik" (Springer-Verlag, Berlin, 1949), p. 63.

(7) APPENDIX

(7.1) Evaluation of the Integral of Eqn. (14)

Integration is carried out along the conductor contours in the ω-plane.

$$P_c = k\int |E(\omega)|^2 \left|\frac{d\omega}{dz}\right| |d\omega|$$

$$|E(\omega)| \simeq \frac{2V}{\pi}$$

where V is the amplitude of the voltage between the strip and the ground planes.

$$\left|\frac{d\omega}{dz}\right| = \frac{\pi}{2h}|\cot\omega|$$

If the integral is taken along the conductor image lines from $v = 0$ to $v = v_p$ we obtain one-fourth of the total losses.

(7.1.1) Losses in the Ground Planes.

Integration is carried out along $u = \pi/2$

$$|\cot\omega| = |\tanh v|$$

$$P_{c1} = 4k\frac{2V^2}{\pi h}\int_0^{v_p} \tanh v\,dv = \frac{V^2 2bk}{h^2} \quad . \quad . \quad . \quad (21)$$

v_p is given by eqn. (5).

(7.1.2) Losses in the Strip.

Integration is carried out along $u = \Delta u$

$$|\cot\omega| = \frac{1}{\sqrt{[(\Delta u)^2 + \tanh^2 v]}}$$

(*a*) $\qquad v = 0 —— n\Delta u \qquad 1 \ll n \ll \dfrac{1}{\Delta u}$

$$P_{c2a} = 4k\frac{2V^2}{\pi h}\int_0^{n\Delta u} \frac{dv}{\sqrt{[(\Delta u)^2 + v^2]}} = \frac{8V^2k}{\pi h}\log 2n \quad . \quad (22)$$

(*b*) $\qquad v = n\Delta u —— v_p$

$$P_{c2b} = 4k\frac{2V^2}{\pi h}\int_{n\Delta u}^{v_p} \frac{dv}{\tanh v}$$

$$= \frac{8V^2k}{\pi h}\left(\frac{\pi b}{4h} - 1 - \log n - \log \Delta u\right) \quad . \quad . \quad (23)$$

v_p given by eqn. (5).

The sum of eqns. (19), (20) and (21) is

$$P_c \simeq \frac{4V^2k}{\pi h}\left(\frac{b}{h} + \frac{2}{\pi}\log\frac{4h}{\pi d}\right)$$

621.372.8

Paper No. 1807 R
July 1955

STANDARD WAVEGUIDES AND COUPLINGS FOR MICROWAVE EQUIPMENT

By A. F. HARVEY, D.Phil., B.Sc.(Eng.), Member.

(*The paper was first received* 16*th July, and in revised form* 10*th December*, 1954.)

SUMMARY

An account is given of the development and standardization of waveguides, couplings and adaptors for microwave equipment. Whilst these items were primarily intended for use in Service electronic equipment, they have proved to be of considerable use in commercial applications. The appropriate sizes of waveguide to cover the whole microwave spectrum are discussed, and reasons are given for the standards chosen. The final specifications are the result of much interchange of information amongst users of equipment in this field, not only in this country but in others. A Section on the finishing and packaging of waveguide parts is included.

(1) GENERAL REQUIREMENTS

(1.1) Microwave Spectrum

Waveguides are used for equipment operated in the microwave part of the spectrum which can be considered to extend from 60 cm to 1 mm wavelength (frequencies from 500 Mc/s to 300 Gc/s). In the early period of work in this field the main applications were on a few selected wavelengths, but in recent years these have extended to new wavebands and also to millimetre wavelengths. Thus when standard sizes of waveguide were being considered it was necessary to choose a comprehensive range which would adequately cover this part of the spectrum.

The choice of particular sizes was governed by a number of technical factors which are discussed in later Sections. Other considerations were the desirability of retaining as many as possible of existing waveguides and the allocation of frequency bands to certain uses—such as the 1 750–2 300 Mc/s, 3 900–4 200 Mc/s and 5 925–8 500 Mc/s bands for civil communication and radio links.[1] At millimetre wavelengths there are a number of regions of high attenuation in the atmosphere which, with the intervening "windows," have a bearing on the wavelengths actually employed.

(1.2) Bandwidth

The total number of waveguide types required to cover the spectrum depends on the bandwidth of transmission, and it is well known that this is limited at the short-wavelength end by either free propagation or insufficient attenuation of higher-order modes, and at the long-wavelength end by increasing attenuation and ultimately complete cut-off. Thus a rectangular waveguide transmitting the dominant H_{01} (TE_{01}) mode has a bandwidth of about 65%, a cylindrical waveguide with the H_{11} (TE_{11}) mode a width of 25% and with the E_{01} (TM_{01}) mode a width of 22%, all approximately centred on a nominal wavelength of two-thirds the cut-off value. These ideal bandwidths are reduced in practice by higher-mode interference, and variations in guide wavelength, impedance and coupling factors. The bandwidth can be increased, however, by using complicated systems which are partly coaxial lines or by the adoption of sections of ridge and dumbbell shapes or with large width/depth ratio. Such considerations of bandwidth also apply to items made partly from waveguide structures such as magnetrons, klystrons, gas switches and aerial systems.

(1.3) Scaling of Designs

It is evident that much development and design effort would be saved if the dimensions of the standard waveguides were directly proportional to their operating wavelengths. Sets of theoretical and experimental data, such as those relating to aerial arrays, rotating joints, directional couplers and waveguide-coaxial transformers could then be applied from one size to another by simple calculation.

Complete scaling of design cannot always be used, since it applies only partially in such cases as crystal-rectifier contacts, cathode emission, properties of materials and power supplies. Moreover, large changes in the dimensions require different manufacturing techniques and operational requirements which would be reflected in design. Nevertheless, the principle of scaling was an important factor in considering a range of standard waveguides.

(1.4) Types of Waveguide

Waveguides serve a variety of functions, and a standard type must be a compromise between varying requirements. Guides of circular section are easy to make, but since the polarization of the dominant mode is not unique, they are used only in special cases where circularly polarized waves, or more than one plane polarized wave, have to be propagated or with H_{01} (TE_{01}) excitation where low attenuation is essential. Waveguides of elliptical section or with semi-circular narrow sides, in which the polarization is defined, are difficult to manufacture.

Where only low powers are used, as in receiving and continuous-wave transmission systems and in test equipment, a compact waveguide is required. This is sometimes achieved by filling the interior with a dielectric such as polystyrene loaded with titanium dioxide which can have a dielectric constant in the range 2·5–10 with a loss tangent of 0·001. Other types of microwave transmission systems include strip lines,[3] in which the guide takes the form of strips of metal foil separated by a dielectric sheet and guiding surfaces.[4]

These transmission systems are regarded as special, and for general purposes rectangular waveguides using the H_{01} (TE_{01}) mode are nearly always used. In the first instance, it is only this type for which standard sizes have been proposed. Simple theory[2] indicates that the width/wavelength ratio should lie between 0·5 : 1 and 1 : 1 and the depth/wavelength ratio should be less than 0·5 : 1, but the practical values to be adopted depend upon the technical factors discussed in Section 2.

(2) RECTANGULAR WAVEGUIDES

(2.1) Optimum Shape and Size

(2.1.1) Power Handling and Attenuation.

The power-handling capacity of a waveguide depends[5] on the cross-sectional dimensions as well as on the breakdown strength of air which, for atmospheric pressure at microwavelengths, is[6] about 30 kV/cm. The net power carried when a mismatched load is present is reduced by a factor S, where

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Harvey is at the Radar Research Establishment.

$S\,(>1)$ is the voltage-standing-wave ratio. It is also reduced by surface irregularities, projections and junctions and varies with electrical parameters such as initial ionization, pulse width and repetition rate. A safety factor of about four in power is normally allowed to cover these points. The power handling can be increased[7] by filling the waveguide with gases such as sulphur hexafluoride and Arcton, or by increasing the pressure, since the power varies at rather less than the square of the pressure.

The attenuation factor due to ohmic losses, at a given wavelength, varies approximately inversely as the guide perimeter.[8] The actual losses are increased in the presence of a standing wave by a factor which is $2(1 + S^2)/(1 + S)^2$ for constant power output from the generator and $(1 + S^2)/2S$ for constant power input to the load. Internal surface roughness[9–11] increases the attenuation factor, and it is necessary to restrict this increase to about 20% by specifying[12] a surface finish of about one-half the skin depth. Typical finishes range from 5 microinches for a silver guide at 3 mm wavelength to 50 microinches for an aluminium guide at 25 cm wavelength.

(2.1.2) Phase Velocity and Impedance.

The phase velocity in a waveguide is greater than that in free space, and it tends to infinity as the operating wavelength approaches cut-off. This variation enhances the phenomenon of frequency splitting[13] occurring under certain conditions when a self-oscillator feeds into a mismatched load at the end of a long transmission line. It is also a disadvantage when two susceptances, such as those of a load and its matching iris, are spaced apart and require cancellation over a band of wavelengths.

When waveguides are coupled together, as in T-junctions, it is desirable that the coupling factor[14] should be as large as practicable, and this usually necessitates a guide of small depth/wavelength ratio. Such a shallow waveguide is also preferred in applications such as linear arrays[15] in which radiating slots are machined in the guide walls. Broad-band matching usually requires that the characteristic impedance be not unduly frequency sensitive. These factors mean that the operating wavelength must not be too near cut-off.

(2.2) Manufacture

It is evident that standard waveguides must be capable of accurate and economic manufacture. In sections greater than $0{\cdot}100\,\text{in} \times 0{\cdot}050\,\text{in}$, waveguides are made by drawing through dies. Practical wall thicknesses and dimensional tolerances range from 0·040 in and 0·001 in respectively for this size, up to 0·125 in and 0·005 in for a section of $6{\cdot}5\,\text{in} \times 3{\cdot}25\,\text{in}$. The

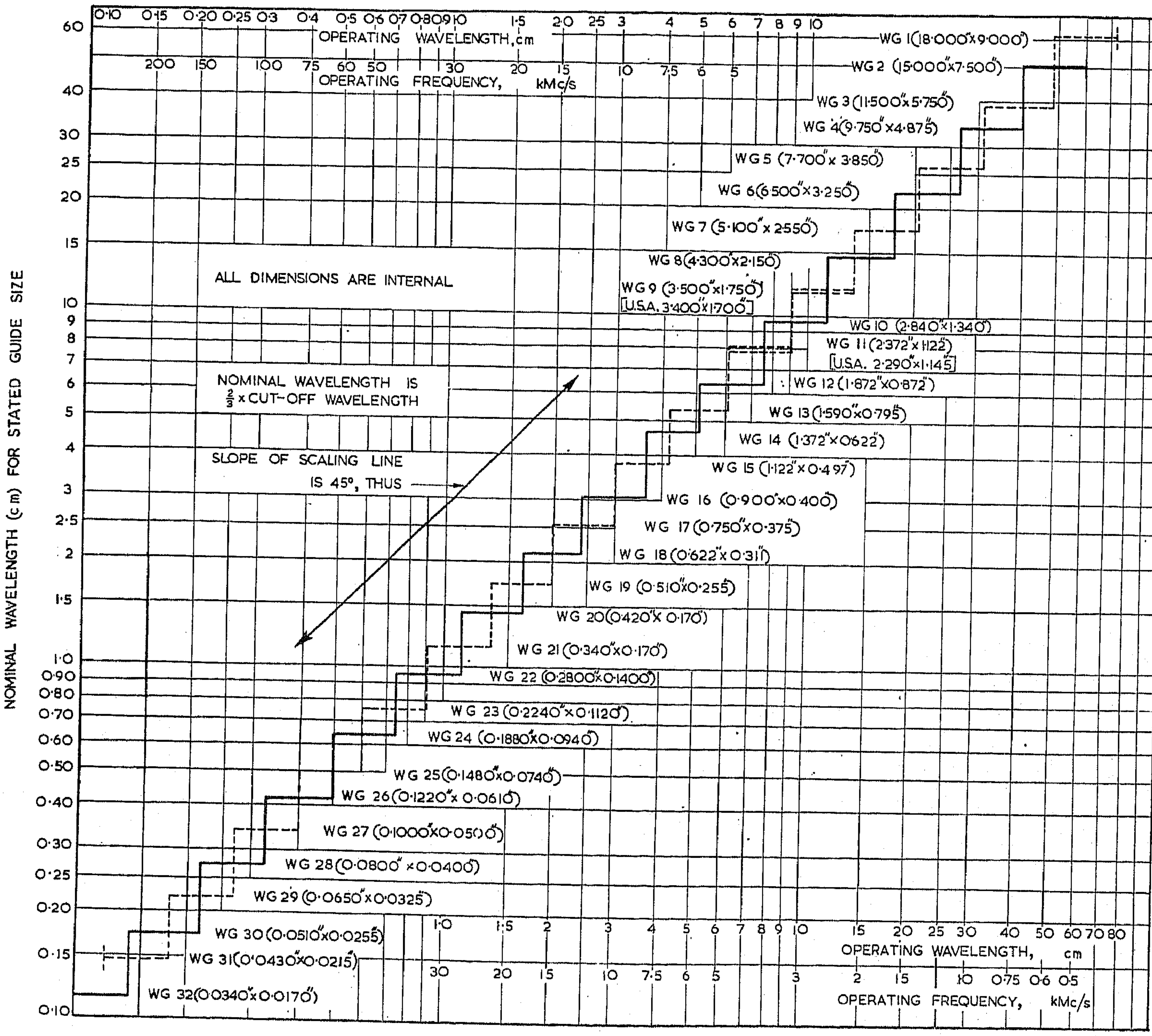


Fig. 1.—Operating ranges of serial rectangular waveguides.

The waveguide dimensions shown are all internal. The nominal wavelength is two-thirds the cut-off wavelength.

Table 1

TECHNICAL DATA FOR STANDARD RECTANGULAR WAVEGUIDES

Guide serial number	Inside rectangle		Cut-off, fundamental mode		Recommended operating range		Attenuation	Power rating
	Width	Depth			Frequency	Wavelength		
	in	in	Gc/s	cm	Gc/s	cm	dB/100ft	kW
WG 1	18·000	9·000	0·327	91·5	0·41–0·61	73·2–48·8	0·0334	103 300
WG 2	15·000	7·500	0·393	76·3	0·51–0·75	60·0–40·0	0·0438	71 800
WG 3	11·500	5·750	0·517	58·2	0·61–0·96	48·8–31·2	0·0614	45 900
WG 4	9·750	4·875	0·590	50·8	0·75–1·12	40·0–26·7	0·0807	31 870
WG 5	7·700	3·850	0·765	39·2	0·96–1·45	31·2–20·7	0·120	18 890
WG 6	6·500	3·250	0·908	33·04	1·12–1·70	26·7–17·7	0·154	13 470
WG 7	5·100	2·550	1·154	25·96	1·45–2·20	20·7–13·6	0·222	8 290
WG 8	4·300	2·150	1·373	21·82	1·70–2·60	17·7–11·5	0·286	5 900
WG 9	3·500	1·750	1·686	17·80	2·20–3·30	13·6–9·10	0·390	3,910
U.S.A.	3·400	1·700	1·733	17·30	2·20–3·30	13·6–9·10	0·400	3 800
WG 10	2·840	1·340	2·080	14·42	2·60–3·95	11·5–7·60	0·555	2 430
WG 11	2·372	1·122	2·485	12·06	3·30–4·90	9·10–6·12	0·726	1 690
U.S.A.	2·290	1·145	2·570	11·70	3·30–4·90	9·10–6·12	0·750	1 600
WG 12	1·872	0·872	3·155	9·51	3·95–5·85	7·60–5·13	1·047	1 040
WG 13	1·590	0·795	3·710	8·08	4·90–7·05	6·12–4·25	1·259	806
WG 14	1·372	0·622	4·285	7·00	5·85–8·20	5·13–3·66	1·700	544
WG 15	1·122	0·497	5·260	5·70	7·05–10·0	4·25–3·00	2·338	355
WG 16	0·900	0·400	6·56	4·57	8·20–12·4	3·66–2·42	3·24	229
WG 17	0·750	0·375	7·87	3·81	10·0–15·0	3·00–2·00	3·92	178
WG 18	0·622	0·311	9·49	3·16	12·4–18·0	2·42–1·67	5·21	123
WG 19	0·510	0·255	11·57	2·592	15·0–22·0	2·00–1·36	7·00	83
WG 20	0·420	0·170	14·08	2·135	18·0–26·5	1·67–1·13	10·9	48
WG 21	0·340	0·170	17·37	1·729	22·0–33·0	1·36–0·91	12·8	37
WG 22	0·280	0·140	21·10	1·423	26·5–40·0	1·13–0·75	17·3	25
WG 23	0·224	0·112	26·35	1·138	33·0–50·0	0·91–0·60	24·0	16
WG 24	0·188	0·094	31·4	0·956	40·0–60·0	0·75–0·50	31·3	10
WG 25	0·148	0·074	39·9	0·752	50·0–75·0	0·60–0·40	44·7	7
WG 26	0·122	0·061	48·4	0·620	60·0–90·0	0·50–0·33	59·7	5
WG 27	0·100	0·050	59·0	0·508	75·0–110	0·40–0·27	80·7	3·0
WG 28	0·0800	0·0400	73·8	0·406	90·0–140	0·33–0·22	113	2·1
WG 29	0·0650	0·0325	90·9	0·330	110–170	0·27–0·18	154	1·4
WG 30	0·0510	0·0255	115·8	0·259	140–220	0·22–0·14	225	0·85
WG 31	0·0430	0·0215	137·5	0·218	170–260	0·18–0·12	286	0·60
WG 32	0·0340	0·0170	173·3	0·173	220–325	0·14–0·09	405	0·37

materials used include copper, brass, silver, occasionally steel, and—in the larger sizes—aluminium. Bimetal waveguides consisting of a brass exterior and a silver lining 0·005in thick are sometimes used. A small radius is allowed on the internal and external corners to give a reasonable life to the drawing dies, while limits are laid down[16,17] for the maximum eccentricity of the internal and external rectangles as well as for permissible bow and twist.

In sizes smaller than 0·280in × 0·140in precision waveguides made by electroforming[18] have come into use, especially where high accuracy of section is required. The external surface is machined cylindrical to simplify the fitting of flanges and give a rigid section.

(2.3) Serial Waveguides

A scheme of serial waveguides covering the microwave spectrum and based on the arguments discussed above was put forward[17] by the author in 1948, and it has since come into general use. The bandwidth chosen is ±10% of the nominal wavelength and the guides fall into two interlocking series shown in Fig. 1. One waveguide can be scaled to another by a line at 45° to the axis, some allowance being made for the varying width/depth ratios caused by the necessity of retaining existing waveguides.

These standard guides have been numbered, and in view of their widespread use their dimensions, cut-off and operating wavelengths (and frequencies), attenuation and power-handling capacity (including safety factor) are given in Table 1. A few of these guides are illustrated in Fig. 2.

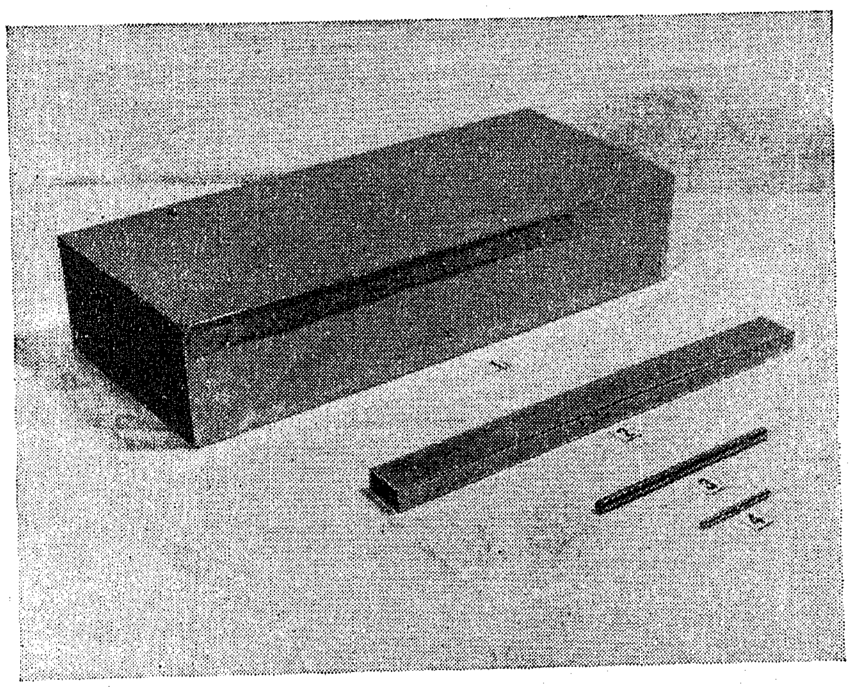


Fig. 2.—Examples of rectangular waveguides.
(1) WG 6 (internal dimensions 6·500in × 3·250in).
(2) WG 16 (0·900in × 0·400in).
(3) WG 24 (0·188in × 0·094in).
(4) WG 28 (0·080in × 0·040in).

Similar schemes have also been proposed in other countries, one being based on two series[19] and another on three series,[20] and they have been of assistance in obtaining international agreement[21,22] on waveguide standards.

(3) WAVEGUIDE COUPLINGS

(3.1) General Design

A waveguide coupling is necessary to hold and align waveguides and consists basically of two flanges, the mating faces of which are flat, smooth and square with the waveguide axis. Satisfactory Service performance also requires that the flanges be identical and symmetrical while there should be no possibility of incorrect orientation. The coupling should be such that any component can be withdrawn laterally from a waveguide system, and provision should be made for sealing the joint against moisture and pressure changes and for preventing power leakage at levels sufficient to cause interference with nearby receivers. The coupling should be capable of use alternatively as a plain contact-shim or choke type, and no loose parts should be present when the flange is initially assembled on its guide.

As an illustration of the dimensional accuracy required, measurements with a lateral misalignment of 2% of the corresponding internal guide dimension, or a rotation of 4°, show a voltage standing-wave ratio (v.s.w.r.) of about 1·005 (0·045 dB*) owing to the susceptance of the discontinuities. If the linear dimensions of the mating guides differed by 0·5% the reflection due to the susceptance of the junction and the change in characteristic impedance would again show a v.s.w.r. of about 1·005 (0·045 dB). Imperfect joints result in contact resistance which produces a small reflection as well as attenuation. The radio-frequency currents in rectangular waveguides propagating the H_{01} (TE_{01}) mode flow longitudinally along the wide faces only, and thus it is important that the standard couplings ensure that the best possible electrical contact is made at these points.

(3.2) Contact-Shim and Choke Types of Coupling

Practical couplings require a v.s.w.r. preferably about 1·002 (0·018 dB), and the corresponding accuracy of alignment is difficult to achieve under Service conditions. A junction less sensitive to small misalignment and imperfections in the mating flanges employs a shim made of soft material, such as aluminium, or a copper-asbestos gasket. In unpublished work at A.S.R.E., an experimental shim for semi-permanent installations has a 0·002 in coating of indium—a soft material which, on clamping, cold-welds to form a good electrical joint.

Another type of shim has a number of spring fingers either twisted or set in opposite directions so that, when the flanges mate, electrical connection is made between the waveguide walls. These shims are not frequency sensitive, and one for WG 11† waveguide was found to have a v.s.w.r. of 1·002 (0·018 dB) over the whole operating wavelength range.

Radio-frequency chokes are another means of making an electrical junction, and although many types have been investigated[23] they usually take the form shown in Fig. 3(*a*) of a coaxial channel or ditch, of depth one-quarter wavelength, spaced radially from the guide walls by an effective quarter-wavelength. Such systems would have a v.s.w.r. not worse than 1·005 (0·04 dB) over a 5% bandwidth. This bandwidth can be improved by making the radial portion of low impedance and the coaxial portion of high impedance, and with a certain empirical choice of dimensions the v.s.w.r. is not worse than 1·02 (0·18 dB) over 20%. Choke junctions are convenient when frequent dismantling is necessary. The preferred arrangement in equipments is to have the plain flange on the generator, then alternatively choke and plain, ending with a choke flange on the load.

With choke systems, misalignment of the waveguide cross-

* The voltage standing-wave ratios mentioned in the paper are defined as greater than unity. Since they are mostly rather small the decibel ratio is given as an alternative.

† For convenience the type numbers only of the guides given in Table 1 are quoted.

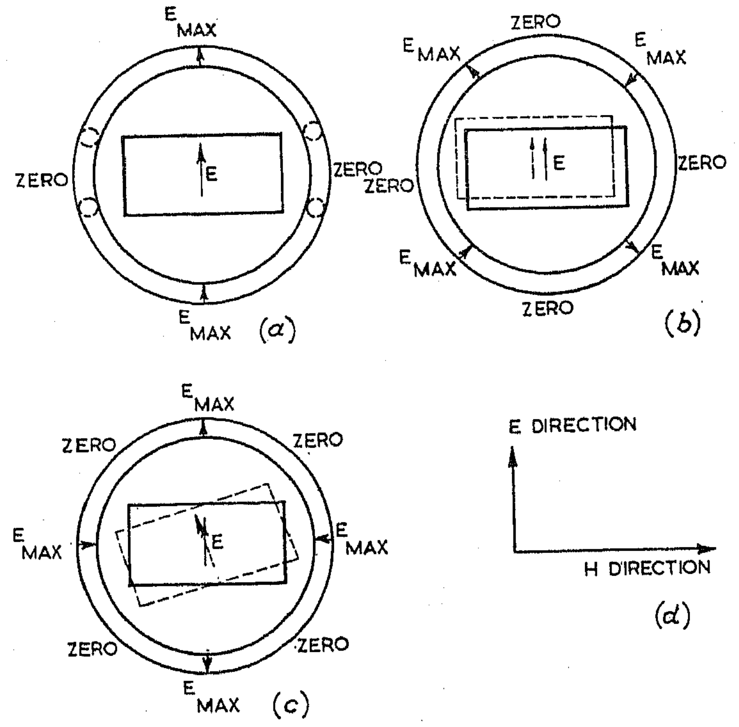


Fig. 3.—Position of the electric field in the waveguide and choke system.

(*a*) With normal excitation.
(*b*) With one orientation of the resonant mode.
(*c*) With the other orientation of the resonant mode.
(*d*) Directions of misalignment (see text).

The misalignments shown are greatly exaggerated and that in any given direction produces a mixture of both resonant modes. The broken circles in (*a*) represent metal slugs which are sometimes used to shift the resonant wavelength.

sections tend to cause resonances, the Q-factor of which may be as high as 500. Probe measurements and observation of arcing show that the second-order mode, with four maxima around the choke periphery, can be excited with the two polarizations shown in Figs. 3(*b*) and 3(*c*). For example, a choke coupling for WG 10 waveguide designed for a 5% band centred on 10 cm wavelength was found to have a v.s.w.r. as large as 10 at 10·75 cm and 10·88 cm wavelength when assembled with exaggerated H-plane [see Fig. 3(*d*)] misalignments of $\frac{5}{32}$ in and $\frac{1}{4}$ in respectively.

A choke for WG 11 waveguide designed for operation from 6·12 to 9·10 cm wavelength gave resonances[24] corresponding to Figs. 3(*b*) and 3(*c*) of 8·53 and 8·46 cm respectively. By the insertion of metal plugs as shown in Fig. 3(*a*) the former resonance was eliminated while the latter was shifted to 7·37 cm wavelength. Some assistance can be obtained by the use of lossy materials positioned so as to damp these unwanted modes, while the alignment and location of the standard couplings is made as accurate as possible.

(3.3) Methods of Sealing

To maintain correct internal pressure and to exclude dirt and moisture the flanges of all standard couplings are provided with grooves for sealing gaskets. The sealing of high-power equipment operating at high altitudes must be adequate at differential test pressures of 50 lb/in^2 over a range of temperature from −50°C to +70°C, while the gasket material must also be proof against attack by oils, mould growth and insects.[25]

These gaskets are of two distinct types. The first, of round section, is made of a material which is capable of distortion but not of compression, and is shown in the coupling illustrated in Fig. 4. The rectangular groove is dimensioned so that, under any condition of tolerances, the gasket is proud by about 15% of its diameter, and on clamping so that the flanges are in full metallic contact, a clearance volume of about 10% remains. The gasket material must not be such as to acquire a permanent

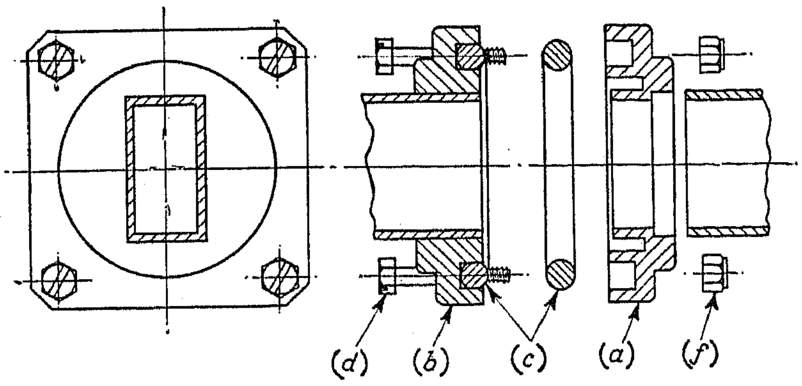


Fig. 4.—Square bolted flange coupling.

The example illustrated employs a broad-band choke flange which is r. :essed to take the waveguide.

(a) Choke flange.
(b) Plain flange.
(c) Gaskets.
(d) Fitted bolt.
(e) Locking nut.

set; synthetic rubbers such as Neoprene to Ministry of Supply specification No. CS2451 are generally used. For operation at temperatures below −30° C natural rubbers to Specifications CS2506 and CS2507, which remain flexible down to −50° C, are

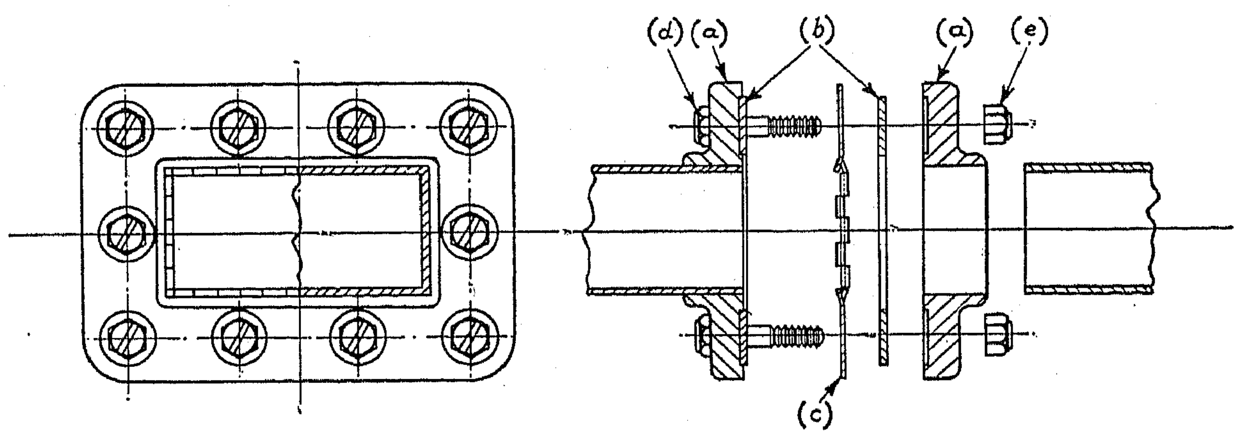


Fig. 5.—Rectangular bolted flange coupling.

The example illustrated uses a spring-finger contact shim.

(a) Plain flange.
(b) Gasket.
(c) Contact shim.
(d) Fitted bolt.
(e) Locking nut.

being considered. When coated with a low-vapour-pressure grease these gaskets have been found satisfactory in continuously-evacuated microwave equipment.

The second type of gasket, of flat rectangular section, is made of a compressible material, and is shown in Fig. 5. The gasket has a thickness about 50% more than the groove depth, and on clamping is compressed to form a seal. After much initial testing a cork–natural-rubber composition, which remains flexible down to −50° C, was chosen as a suitable material. It has a Shore hardness factor of 70, and damage by oil and insects is reduced by the protection of the groove walls, and that due to mould growth, by impregnation with a fungicide.[26]

(3.4) Fitting and Clamping Arrangements

As a method of fitting coupling flanges to waveguides the simple butt joint has not come into general use. Improved strength and alignment can be obtained by a socket fitting in which the flange has a recessed portion to take the waveguide. A choke flange of this type is shown in Fig. 4, the corresponding plain flange having a through type of fitting. The latter method is used where possible, although care has to be taken in assembly to prevent damage to the mating face of the flange. Yet another fitting is shown in Fig. 6, in which the flange has a cylindrical hole to accept precision electroformed waveguide.

The standard flanges are designed for medium-temperature brazing such as by the Easiflo process to give a joint which is strong and fatigue resistant. For millimetre-wavelength sizes, lightly stressed components supported inside equipment, and for accurate test-gear soft solders are desirable to avoid the distortion caused by high temperatures.

The design of flanges in aluminium and similar light materials has presented difficulty, since existing methods of attachment are not entirely satisfactory. In one method the areas of contact are electroplated with copper and then joined with soft solder. Such a joint would rapidly fail under humid or salt-spray conditions, but one sample, after coating with Araldite[27] resin, cured at 130° C for 7 hours, passed satisfactorily all the relevant tests of Specification K114.[28] This resin can also be used as a jointing adhesive provided that some mechanical pinning is incorporated to give added strength. Various forms of vibratory soldering irons[29] can be used to detach the oxide film which forms on aluminium to enable tinning and subsequent soft soldering to take place.

A number of proprietary aluminium soft-solders have been tried, one product with a working temperature of 200° C showing slight corrosion after humidity testing, while another, working at 450° C, gave a better performance. The combination of metals present in these medium-temperature processes tends to cause corrosion and prevent anodizing, although protection can again be given by Araldite.

Stronger and more reliable joints can be made by methods using pure metals and non-corrosive fluxes. In dip brazing the parts, which must be almost pure aluminium, are jigged with a

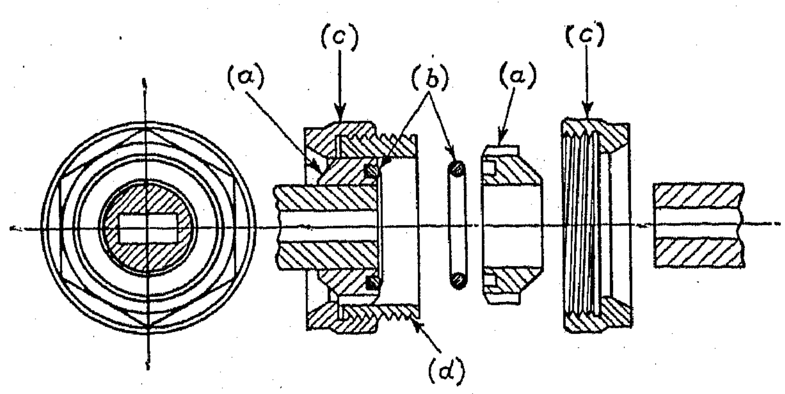


Fig. 6.—Screwed ring coupling.

The example illustrated has plain lapped faces and is for use with WG 22 precision waveguide.

(a) Plain flange.
(b) Gasket.
(c) Nut ring.
(d) Locating ring.

filler material containing an alloy rich in aluminium. The assembly is then immersed in a mixture of salts which acts as a flux and raises the temperature to a point in excess of the melting point of the filler but below that of the aluminium. A flame-brazing method works at a temperature of 580° C, but since this is only some 40° C below the melting point of aluminium the operation is a skilled one. Joints made by these and also argon-arc methods can, after anodizing, withstand satisfactorily the tests of Specification K114. Standard couplings are made from grades of aluminium suitable for use with these processes.

A simple but satisfactory means of holding and aligning waveguide flanges is by close-tolerance bolts fitting in accurately bored and positioned holes. Self-locking nuts are used, and to conform with international standards the new unified thread system[30–32] is used. The coupling illustrated in Fig. 4 is suitable for either plain or choke-type connection for waveguides WG 10 to WG 20 and is sealed by round-section gaskets. Plain rectangular flanges are more suitable with larger waveguides, and such an arrangement, using a flat-section gasket and contact shim, is shown in Fig. 5. In this coupling the highest accuracy of alignment is ensured by drilling two of the flange holes after assembly by means of a jig located from inside the waveguide.

When the coupling has to be frequently disengaged the more convenient form of clamping shown in Fig. 6 is employed, in which lateral and angular alignment of the flanges is provided by a keyed cylindrical sleeve. The flanges are provided with auxiliary flats so that after assembly on the guide the clamping-ring nut can be slipped over them.

(4) ADAPTORS

Adaptors are required to connect flanges of different type or design, such as those of a different country.[33] When the mating flanges incorporate chokes their series reactances are made to cancel by making the electrical length of the adaptor an odd number of quarter-wavelengths. Such an adaptor on guide WG 11 using broad-band chokes was found to have v.s.w.r.'s of 1·019 (0·171 dB), 1·012 (0·108 dB), 1·010 (0·090 dB) and 1·012 (0·108 dB) at 9·38, 8·83, 8·33 and 8·22 cm wavelengths respectively.

Another adaptor would incorporate an impedance-matching section, such as a half-wavelength taper, to couple waveguides of different section. The mean guide wavelength in such a taper is empirically equal to one-sixth of the wavelength in a guide with dimensions equal to the input section, plus two-thirds of that of the section half-way along, plus one-sixth of that of the output section.

Adaptors in the larger sizes are made by brazing flanges to the appropriate length of guide, while those in the smaller sizes are more suitably constructed by electroforming techniques,[14] the flanges, in some cases, being separately made and soldered on to the guide. Medium-size adaptors can be made in one

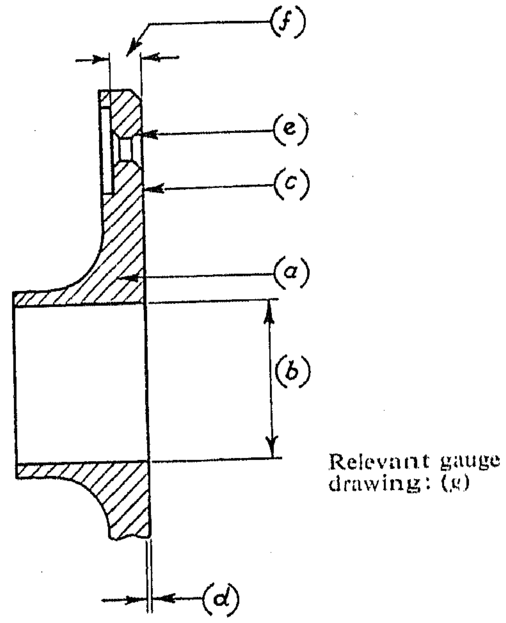


Fig. 8.—Typical parameters requested on the drawing of a waveguide part.

(a) Material.
(b) Dimensions with tolerances.
(c) Surface finish.
(d) Flatness and squareness tolerances.
(e) Chamfers on sharp edges of mating faces.
(f) Spot facing of holes to specified dimensions.
(g) Use of gauges for mechanical inspection.

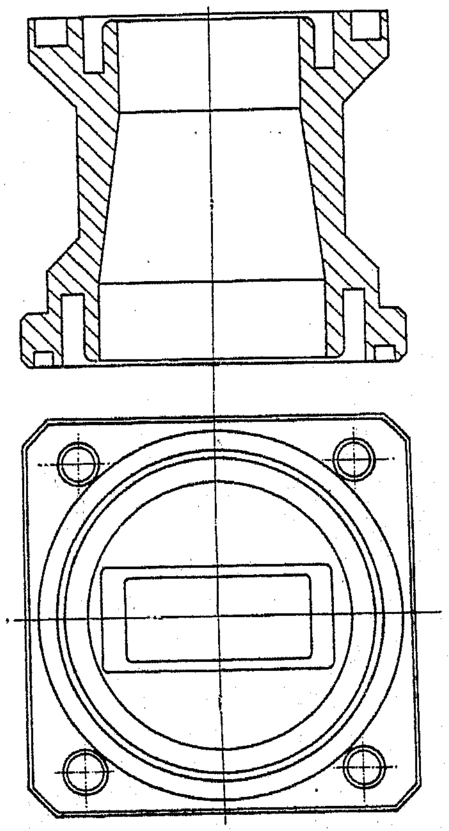


Fig. 7.—Conversion adaptor.

The example illustrated is made in one piece from light-alloy material and connects together WG 15 and WG 16 waveguides. The electrical length of the taper section is one-half wavelength and that between the choke flanges is one wavelength.

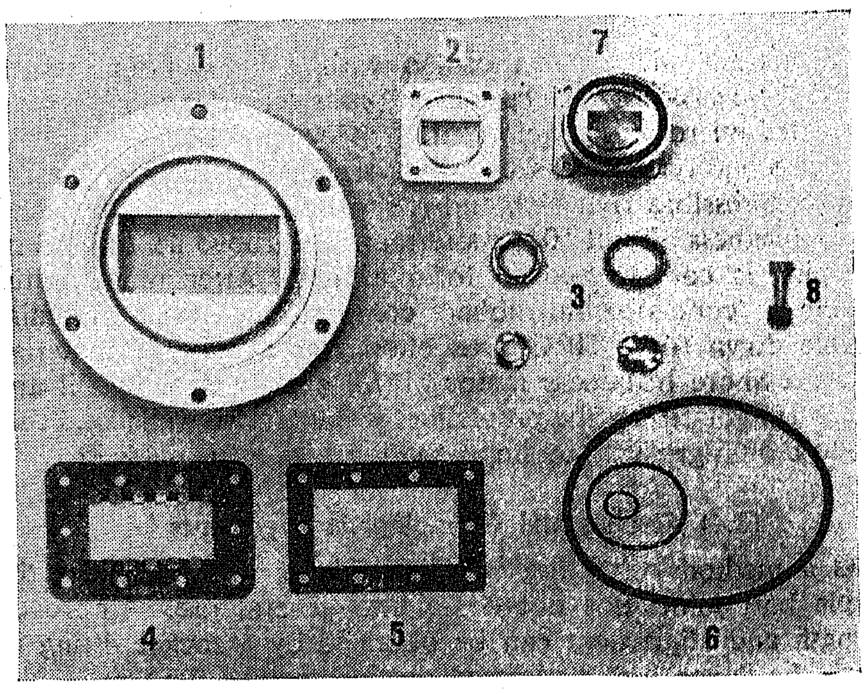


Fig. 9.—Parts of waveguide couplings and adaptors.

(1) Bolted choke flange for WG 10 guide.
(2) Bolted choke flange for WG 15 guide.
(3) Screwed ring coupling for WG 22 guide.
(4) Contact shim for WG 11 guide.
(5) Flat cork-rubber gasket for WG 11 guide.
(6) Round rubber gaskets for WG 10, WG 16 and WG 22 guides.
(7) Adaptor from WG 15 to WG 16 guide.
(8) ¼ in diameter fitted bolt and locking nut.

piece as a simple microwave component. Suitable manufacturing methods would include pressure die casting, materials such as Mazak being satisfactory, provided that the zinc is pure enough to give corrosion-resistant properties.[34] Light alloys present difficulty but are amenable to such precision casting processes as lost wax and Mercast.[35] The method employed for the light-alloy taper from WG 15 to WG 16 shown in Fig. 7 is hobbing, in which a tool, machined and ground to the shape of the internal section required, is pressed into a metal blank and then withdrawn.

Mention might be made here of the specifications and drawings[36,37] of these couplings and adaptors in view of the difficulty of stating in engineering terms the manufacturing requirements for waveguide parts. Reliance has often to be placed on the acquisition of special techniques and microwave testing,[38] but for Service components which must be produced economically in large quantities, held in stores and finally assembled in remote situations, such methods are impracticable. There has been much investigation of means of putting the electrical requirements into terms which can be specified on production drawings. Fig. 8, for example, specifies parameters such as material, dimensions and tolerances, surface finish,[39] flatness and squareness tolerance,[40] the provision of chamfers on sharp edges, spot facing of holes and the use of mechanical gauges. Some examples of these standard couplings and adaptors in the final form are shown in Fig. 9.

(5) FINISHING OF WAVEGUIDE PARTS

(5.1) Protective Finishes

Provided that the sealing facilities of the waveguide couplings are used in conjunction with proper desiccator breathing systems, the inside surfaces of waveguide equipment require only slight protection. Varnishes such as cellulose nitrate, Specification X17, seaplane or beeswax dissolved in benzene have been found satisfactory. The open ends of the waveguide system, usually at the aerial, can be sealed with thin polythene tape.

External finishes should be chosen to give electrolytic potential differences with neighbouring metals not greater[26] than 0·5 volt for ordinary purposes and 0·25 volt when exposure to the weather and salt spray is likely. Nickel or cadmium plating has been found satisfactory for parts constructed from other than light alloys, for which anodization is recommended. For small laboratory instruments, plating with 10^{-4}in of silver and 2×10^{-6}in of rhodium gives a hard wearing and attractive finish which has good resistance to humid conditions. Corrosion inhibitors[41] may also have applications.

The variable thickness of commercial electroplating is not satisfactory for closely fitting parts such as locating bolts and sleeves, and these are made from corrosion-resistant alloys such as stainless steel, Tungum and nickel-silver. Screw threads, especially in stainless steel and light alloys, should be coated before assembly with graphite grease to reduce liability to pick up after prolonged use. Contact shims are made of beryllium-copper, and throughout the standard ranges of waveguide parts discussed, the design and materials have been chosen to withstand the full relevant Service specification.[28]

(5.2) Packaging and Marking

Special care should be taken to ensure that such precision items as waveguide components are packed so as to afford protection from damage and distortion during storage, transit and handling. For example, waveguide tubing for Service stores is packed so that each length is in a separate cleated crate from which the guide is removed only when required for use. Large components such as flanges and adaptors can be packed in strong cartons, while small items such as millimetre-wavelength guides, fitted bolts and nuts should additionally be enclosed in sealed polythene bags. Waveguide test-gear is usually supplied by the manufacturers in fitted wooden cases, in which they should be replaced when not in use.

The ends of the waveguide can be sealed with plastic inserts, and these are sometimes provided with extensions to protect the mating faces of the flanges. During investigation of sealing methods a 24in length of WG 16 guide, containing a small quantity of silica-gel desiccant and sealed with one standard type of moulded-polythene plug, was subjected to cycling between +25°C and +35°C in an atmosphere of 95% relative humidity. After 14 days there was only a slight loss of brilliance of the internal surface and an increase in weight of the desiccant, due to moisture absorption, of only 0·4gm.

Care should be taken in identification marking so that no damage to the electrical, mechanical and sealing properties of the equipment occurs. Light stencil marks or transfers in positions fulfilling no electrical or essential mechanical function are permissible. Transit or packing cases should also be marked or labelled.

(6) ACKNOWLEDGMENTS

The author wishes to acknowledge assistance received from the Engineering Department of the Radio Research Establishment, whose work has included mechanical and tropical tests on the various items and the preparation of production drawings. He is also indebted to Evered and Co., Ltd., and his colleagues on the Radio Components Research and Development Committee and the Radio Components Standardization Waveguide Committee for much useful discussion. Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller, H.M. Stationery Office, for permission to publish the paper.

(7) REFERENCES

(1) "International Radio Regulations," Atlantic City Convention, 1947.

(2) Ragan, G. L.: "Microwave Transmission Circuits" (McGraw-Hill Book Co., New York, 1948).

(3) Grieg, D. D., and Engelman, H. F.: "Microstrip—A New Transmission Technique for the Kilo-Megacycle Range," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1644.

(4) Goubau, G.: "Surface Waves and their Applications to Transmission Lines," *Journal of Applied Physics*, 1950, **21**, p. 1119.

(5) Barlow, H. M.: "The Relative Power-carrying Capacity of High-Frequency Waveguides," *Proceedings I.E.E.*, Paper No. 1225 R, January, 1952 (**99**, Part III, p. 21).

(6) Cooper, R.: "Experiments on the Electric Strength of Air at Centimetre Wavelengths," *Journal I.E.E.*, 1947, **94**, Part III, p. 315.

(7) Sutherland, J. W.: "A Study of Electrical Breakdown in Waveguides at 3000 Mc/s," *Electronic Engineering*, 1954, **26**, p. 538.

(8) Kuhn, S.: "Calculation of Attenuation in Waveguides," *Journal I.E.E.*, 1926, **93**, Part IIIA, p. 663.

(9) Benson, F. A.: "Waveguide Attenuation and its Correlation with Surface Roughness," *Proceedings I.E.E.*, Paper No. 1467 R, March, 1953 (**100**, Part III, p. 85).

(10) Beck, A. C., and Dawson, R. W.: "Conductivity Measurements at Microwave Frequencies," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 1181.

(11) MORGAN, S. P.: "Effect of Surface Roughness of Eddy Current Losses at Microwave Frequencies," *Journal of Applied Physics*, 1949, **20**, p. 352.

(12) HARVEY, A. F.: "A Surface-Texture Comparator for Microwave Structures," *Proceedings I.E.E.*, Paper No. 1756 R, March, 1955 (**102** B, p. 219).

(13) COLLINS, G. G.: "Microwave Magnetrons" (McGraw-Hill Book Co., New York, 1948).

(14) CALLICK, E. B.: "Design for Compensated T-Junction in Rectangular Waveguide" (unpublished Ministry of Supply Report).

(15) FRY, D. W., and GOWARD, F. K.: "Aerials for Centimetre Wavelengths" (Cambridge University Press).

(16) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "Specification for Waveguides, Seamless Rectangular Tubing RCS 351."

(17) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "List of Standard Waveguides, Seamless Rectangular Tubing RCL. 351."

(18) HARVEY, A. F.: "The Electroforming of Components and Instruments for Millimetre Wavelengths," *Proceedings I.E.E.*, Paper No. 1757 R, March, 1955 (**102** B, p. 223).

(19) FENN, W. H.: "Standards for Waveguides," *Electronics*, June, 1949, p. **110**.

(20) "Guides d'Ondes Rectangulaires": Specification C.C.T.U., 317, December, 1950.

(21) "Military Specification—Waveguides, Rigid Rectangular MIL-W-85C," Armed Services Electro Standards Agency, Fort Monmouth, United States.

(22) Joint Canadian Specification: "Tubing, Waveguide, Seamless Rectangular JCNAAF-T-18," Department of National Defence, Ottawa.

(23) HARVEY, A. F.: "Instruments for use in the Microwave Band," *Proceedings I.E.E.*, Paper No. 1156 M, April, 1951 (**98**, Part II, p. 781).

(24) DAVIES, D. J.: "Resonance Effects in Choke-Type Connectors for use in No. 11 size Waveguide" (unpublished Ministry of Supply Report).

(25) GIBLIN, J. F., and KING, W. T.: "The Damage to Lead-sheathed Cables by Rodents and Insects," *Proceedings I.E.E.*, Paper No. 1641 S, May, 1954 (**101**, Part I, p. 123).

(26) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "General Requirements for Service Telecommunication Equipment," RCS. 1000.

(27) MOSS, C. J.: "Araldite—A New Adhesive, Coating and Casting Resin," *British Plastics*, 1948, **20**, p. 521.

(28) WIRELESS TELEGRAPHY BOARD SPECIFICATION: "Climatic and Durability Testing of Service Telecommunications Equipment," K.114.

(29) NOLTINGK, B. E., and NEPPIRAS, E. A.: "Ultrasonic Soldering Irons," *Journal of Scientific Instruments*, 1951, **28**, p. 50.

(30) "Unified Screw Threads," B.S. 1580: 1953.

(31) "Unified Precision Hexagonal Bolts, Screws, Nuts and Plain Washers—Normal Series," B.S. 1768: 1951.

(32) "Unified Machine Screws and Machine Screw Nuts," B.S. 1981: 1953.

(33) Military Specification: "Flanges and Associated Fittings for Rectangular Waveguide Couplings—MIL-F" (Armed Services Electro Standards Agency, Fort Monmouth, United States).

(34) PENNY, F. D.: "A Handbook of Die Castings," H.M. Stationery Office, 1953.

(35) NEIMEYER, W. I.: "Precision Casting with Frozen Mercury Patterns," *Iron Age*, 17th March, 1949.

(36) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "Specification for Standard Waveguide Couplings," RCS352.

(37) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "List of Standard Waveguide Couplings," RCL.352.

(38) BROWN, L. W.: "Production Testing of Waveguides," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 47.

(39) "The Assessment of Surface Texture," B.S. 1134: 1950.

(40) "Engineering Drawing Practice," B.S. 308: 1953.

(41) PRESTON, J. N.: "Corrosion Control with Organic Inhibitors," *Industrial and Engineering Chemistry*, 1952, **44**, p. 1755.

621.372.8 : 621.3.011.21

Paper No. 1882 R
July 1955

ON THE SURFACE IMPEDANCE OF A CORRUGATED WAVEGUIDE

By A. E. KARBOWIAK, Ph.D., B.Sc.(Eng.).

(*The paper was first received* 6*th January, and in revised form* 31*st March*, 1955.)

SUMMARY

An examination of experimental results on corrugated surface waveguides as reported recently by Barlow and Karbowiak is carried out. An empirical modification to the theoretical formula for the surface impedance is derived which is applicable for all values of surface parameters, provided that the pitch of corrugation is less than about one-quarter of the wavelength.

LIST OF PRINCIPAL SYMBOLS

μ_0, ϵ_0 = Permeability and permittivity* of free space (in rationalized M.K.S. system of units).
f = Frequency.
$\omega = 2\pi f$ = Angular frequency.
$K = j\omega\sqrt{(\mu_0\epsilon_0)} = j\dfrac{2\pi}{\lambda_0}$ = Free-space propagation coefficient.
λ_0 = Free-space wavelength.
Z_s = Surface impedance.
X_s = Surface reactance.
X_m = Measured value of surface reactance.
X_t = "Theoretical" value of surface reactance, defined by eqn. (3).
s = Guide radius (outside).
$s' = s - l$ = Guide radius (core).
l, d, D = Parameters of a corrugated surface (see Fig. 1).

(1) INTRODUCTION

In a recent paper[1] a series of experiments on cylindrical surface waves supported by corrugated guides have been described. The purpose of the experiments was an empirical determination of the surface reactance of corrugated surfaces as a function of the surface parameters and frequency. The results of the experiments were collected in the form of a number of graphs and the authors pointed out, at the time, the lack of agreement between the experimental results and present theories, unless the depth of the corrugations is substantial.

The present paper reviews the results of the previous paper[1] with a view to establishing an empirical formula for the surface reactance of a corrugated surface.

The quantity under investigation, X_s, is the imaginary part of the surface impedance $Z_s = R_s + jX_s$, which is defined by

$$Z_s = \left.\frac{E_t}{H_t}\right|_S \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where the suffix t denotes the tangential components (to the surface of the waveguide, S) of the field vectors. With a corrugated surface (an example of an anisotropic surface) the value of Z_s depends on the orientation of the E_t-vector, and in general

$$Z_s = [Z_\phi, Z_x] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where Z_ϕ and Z_x are the circumferential and axial "components" of Z_s. In what follows the component Z_x is implied.

(2) DISCUSSION OF PUBLISHED RESULTS PERTAINING TO THE SURFACE IMPEDANCE OF A CORRUGATED SURFACE

Referring to the results published elsewhere,[1] the theoretical formula for the surface reactance of a metal waveguide of proportions as indicated in Fig. 1 is given by

$$X_t = Z_{slot} \times \frac{d}{D} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where Z_{slot} is the TEM impedance presented by the slot (which

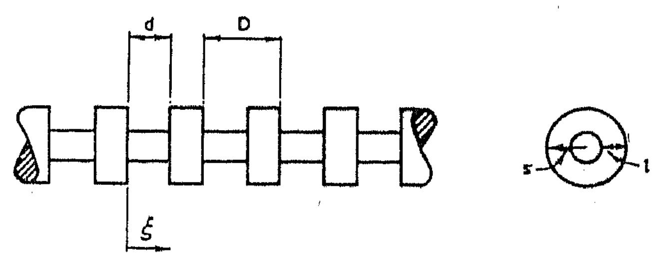


Fig. 1.—Corrugated guide.

is in fact a short-circuited piece of radial transmission line) and is given by

$$Z_{slot} = -jZ_0\frac{Y_0(-jKs')J_0(-jKs) - J_0(-jKs')Y_0(-jKs)}{Y_0(-jKs')J_1(-jKs) - J_0(-jKs')Y_1(-jKs)} \quad . \quad (4)$$

In the derivation of eqn. (3) the conductor out of which the waveguide is made has been assumed to be of infinite conductivity, and furthermore eqn. (3) is only a first-order approximation valid under the assumptions that

$$\left.\begin{array}{r} D \ll \lambda_0 \\ \text{and} \qquad d \ll l \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

It is thus not surprising that the experimental results were in agreement with eqn. (3) only when expressions (5) were fulfilled, and that an equation for the surface reactance of a corrugated surface for which expressions (5) are not satisfied is still required.

(3) A GENERAL FORMULA

Although the utility of eqn. (3) is restricted by the assumptions in expressions (5), it is possible, using published experimental results,[1] to derive an empirical relation which is free of these restrictions, in the following manner.

Let us postulate that the surface reactance, X_s, of a corrugated surface is given [instead of by eqn. (3)] by

$$X_s = X_t f(l/d) \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

* *Note.*—The paper is a sequel to "An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides," by H. E. M. Barlow and A. E. Karbowiak, *Proceedings I.E.E.*, Paper No. 1625 R, May, 1954 (101, Part III, p. 182) in which the symbol κ denoted permittivity.
If the two papers are read in conjunction, therefore, it should be borne in mind that the symbol κ in the earlier paper has the same significance as the symbol ϵ (B.S. 1991 : 1954) in the present paper.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Karbowiak is at Standard Telecommunication Laboratories Ltd.

where $f(l/d)$ is some function (as yet unknown) of l/d and X_t is given by eqn. (3). Now let X_m be the measured value of surface reactance. If eqn. (6) is correct, all experimental results when expressed as X_m/X_t and plotted against l/d would have to lie on a common curve [since $X_m/X_t = f(l/d)$] no matter how diverse the conditions of measurements may have been. This, indeed, would be the function $f(l/d)$ characteristic of all corrugated surfaces whose corrugations are rectangular.

Fig. 2 is a plot of the quantity X_m/X_t using experimental

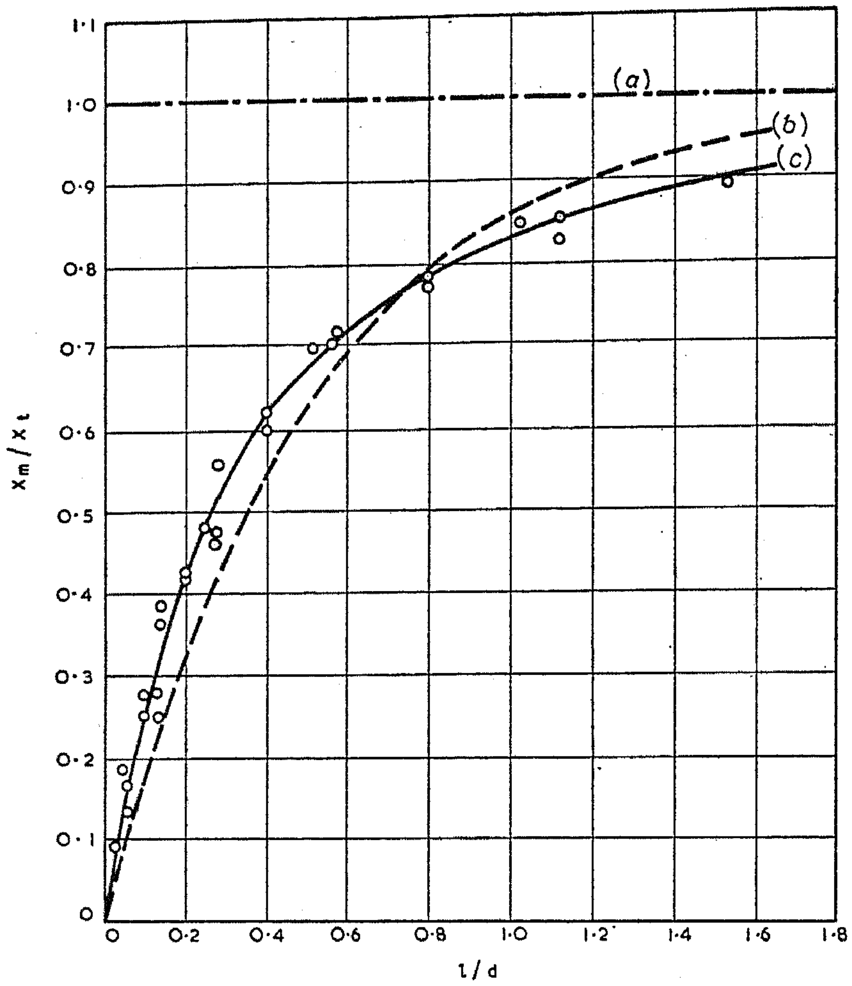


Fig. 2.—X_m/X_t as a function of l/d.

results published previously and the points shown in the Figure were taken at random from the experimental curves.[1] It is evident that the experimental points cluster about the curve (*c*); further, the scatter of points is totally uncorrelated and is ascribed principally to the experimental errors. This, then, demonstrates the existence of a single function, $f(l/d)$, peculiar to a corrugated surface whose corrugations are rectangular.

Before elaborating on our findings it is worth examining the general behaviour of the $f(l/d)$ curve: for large values of its argument it tends towards unity, and the surface reactance is then given by eqn. (3) to an adequate degree of approximation [curve (*a*)].

It is not difficult to show analytically that the equation to curve (*c*) is given by

$$f(x) = 1 - \tfrac{1}{2}\varepsilon^{-5x} - \tfrac{1}{2}\varepsilon^{-x} \quad . \quad . \quad . \quad . \quad (7)$$

where

$$x = l/d \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus we conclude that the surface reactance of a corrugated surface whose corrugations are rectangular is given by

$$X_s = X_{slot}(d/D)(1 - \tfrac{1}{2}\varepsilon^{-5x} - \tfrac{1}{2}\varepsilon^{-x}) \quad . \quad . \quad . \quad (9)$$

where X_{slot} is given by eqn. (4).

Eqn. (9) is valid at any frequency and for any proportions of the corrugations with one provision only, namely that it does not apply where D, the pitch of corrugation, is comparable with λ_0. In other words, eqn. (9) cannot account for the "recovery effect" as reported by Barlow and Karbowiak [see Fig. 11 of their paper],[1] since this occurs for $D > \lambda_0/3$; consequently, eqn. (9) is applicable subject to the condition $D < \lambda_0/4$.

Incidentally, if the corrugations are very shallow [say $l < \lambda_0/10$] we can show—by expanding the various Bessel functions, occurring in eqn. (4), in Taylor's series about the point $(-jKs)$, and by retaining the first term of the respective expansions only—that

$$Z_{slot} = Z_0 k_0 l$$

$$= 2\pi\left(\frac{l}{\lambda_0}\right)Z_0 \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Similarly, if l is small we can approximate to the exponentials occurring in eqn. (7) by the first two terms of their series, and we then get

$$f\left(\frac{l}{d}\right) \simeq \left[1 - \frac{1}{2} + \frac{5}{2}\left(\frac{l}{d}\right) - \frac{1}{2} + \frac{1}{2}\frac{l}{d}\right]$$

$$= 3\frac{l}{d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Substituting eqns. (10) and (11) into eqn. (6) we get

$$X_s \simeq 2\pi Z_0\left(\frac{l}{\lambda_0}\right)\frac{d}{D}3\frac{l}{d} \quad . \quad . \quad . \quad . \quad (12)$$

i.e.

$$X_s = 6\pi Z_0\frac{l^2}{\lambda_0 D} \quad (l \text{ being small})$$

This, then, establishes the law

$$X_s \propto l^2 \quad (l \text{ being small}) \quad . \quad . \quad . \quad . \quad (13)$$

which has been shown (experimentally) to be the case [see Fig. 15 of Reference 1].

Passing now to the other extreme of large values of l/d, we have

$$f(l/d) \to 1 \quad (l \text{ being large}) \quad . \quad . \quad . \quad . \quad (14)$$

and consequently

$$X_s \to X_{slot}(d/D) \quad (l \text{ being large}) \quad . \quad . \quad . \quad (15)$$

It is of some practical interest to note that curve (*b*), whose equation is of much simpler form, namely

$$f(l/d) = 1 - \varepsilon^{-2(l/d)} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

is already a good approximation (within a few per cent) to the actual law.

(4) ACKNOWLEDGMENTS

Acknowledgment is made to Standard Telecommunication Laboratories Ltd. for the facilities offered in the preparation of the manuscript and for permission to publish the paper.

(5) REFERENCES

(1) Barlow, H. M., and Karbowiak, A. E.: "An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides," *Proceedings I.E.E.*, Paper No. 1625 R, May, 1954 (**101**, Part III, p. 182).

621.396.11 : 551.510.535

Paper No. 1855 R
July 1955

AN IMPROVED CHART FOR IONOSPHERIC FORECASTING IN THE BRITISH ZONE

By R. NAISMITH, Member.

(*The paper was first received* 28*th January, and in revised form* 5*th March,* 1955.)

SUMMARY

An improved form of chart for use in ionospheric forecasting is described. In the international 3-zone system of presenting forecasts of radio-wave transmission conditions, the necessity to provide average values over a range of longitudes introduces errors. These errors can be eliminated for the region of the United Kingdom by the use of a British Zone Chart (B.Z.C.), which presents the predictions directly for all 2000 km control points around London.

A comparison by various user organizations showed that, over a test period of six months, the standard deviation between the predicted and observed conditions was approximately halved when the B.Z.C. was used.

(1) INTRODUCTION

The ionospheric conditions which govern the propagation of radio waves over the world are difficult to portray on any simple form of chart. This is primarily due to the fact that four variables have to be considered in any method of presentation, namely latitude, longitude, frequency and time. In the 3-zone system of presentation which has been adopted internationally for a number of years, the parameters are latitude, frequency and time. The world is then divided into two zones each covering 120° of longitude, and it is assumed that over this limited range of longitudes the same diurnal variation of frequency occurs. It is this assumption which gives rise to considerable errors, and they are particularly high near longitudes of 0° and 180°, because the values for these two extreme longitudes are averaged and plotted on a third zone chart. It was anticipated that this would involve a substantial error, and in consequence a Table giving the magnitude of this error has been issued with each D.S.I.R. prediction of radio-wave communication conditions since the 3-zone system was adopted in 1946.

The United Kingdom is centred approximately on longitude 0°, and one of the control points for all circuits to and from this country is situated in or on the border of this region of low accuracy. It is also probable that there is a greater density of radiocommunication in this country than in any other area of similar extent. This is largely due to the method by which the British Commonwealth operates as an integral unit. Actually, the area on the charts corresponding to the North Pole, where there is little interest for normal radiocommunication, is greater than the area on the charts corresponding to this United Kingdom region of great density of radio circuits. There are thus good reasons to warrant an increase in the accuracy of predictions for the United Kingdom area.

(2) BRITISH ZONE CHARTS

This increased accuracy was made available in the form of a British Zone Chart (B.Z.C.), for an experimental period of 12 months, on which the predictions for all points lying on the periphery of a 2000 km circle centred on London were presented. It was convenient to give the predictions in the form of the 4000 km maximum usable frequencies (m.u.f.'s), because these could be read directly for any circuit to or from the United Kingdom when it was controlled by the ionospheric conditions at the United Kingdom end. The chart was plotted on a scale of latitude and Universal Time (U.T.). Since at any latitude there are two possible control points, one to the east and the other to the west, the chart was a composite one. One-half of the chart referred to circuits to and from the west and the other half to circuits to and from the east. One example of this type of chart is shown in Fig. 1. The scales can and have been altered as required. It will be noted that, in addition to the increased inherent accuracy, it is also in the form of a direct-reading chart. For example, the names of various circuits from London have been written at the latitude of the control point at the London end, and it will be seen that, not only can the 24 hourly values for the circuit be read off directly, but values at intermediate times are also available; there is, in fact, no discontinuity in time. This type of chart was issued with predictions of radio-wave communication conditions for each month from June, 1953, to May, 1954.

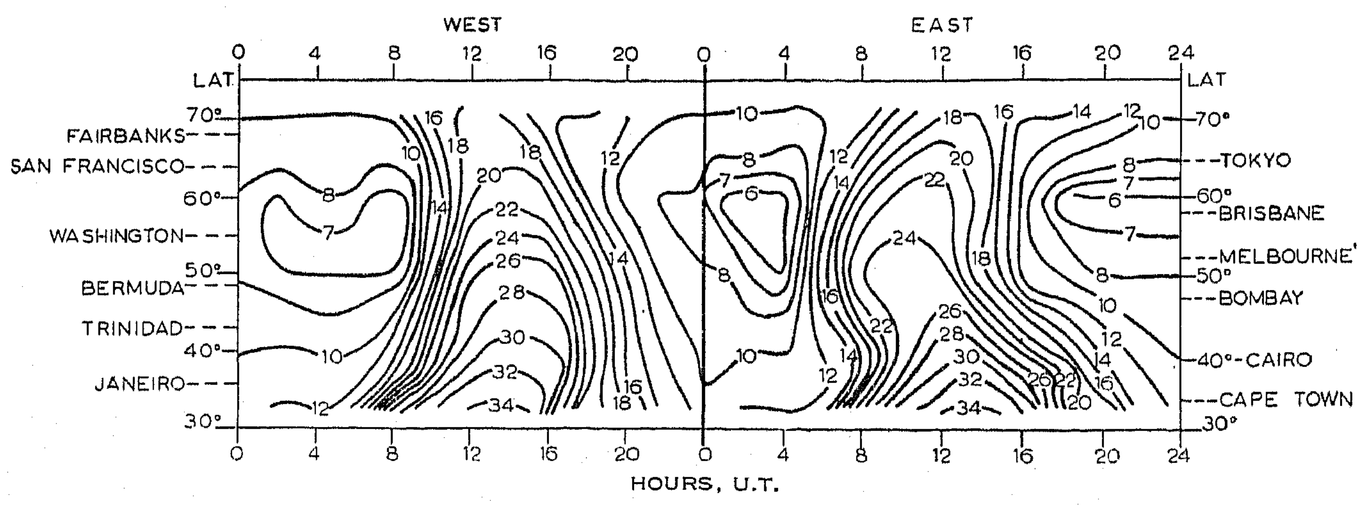


Fig. 1.—British Zone Chart.

4000 km m.u.f. contours (in megacycles per second) for December, 1950.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

(3) INCREASED ACCURACY OF BRITISH ZONE CHARTS

The error in the 3-zone chart system, which can be eliminated when the British Zone Chart is used, may be divided into two parts; one is due to the averaging process mentioned above, and the other to the assumption that time and longitude may be interchanged over the 60° of longitude involved in the B.Z.C. The standard deviation computed for the first error is 16%, and for the latter, 8%. There is also a gain associated with the larger scale adopted for the B.Z.C. This is due to the fact that, while the same size of chart is used, the actual area to which it refers is only one-fortieth of that of the 3-zone chart.

It should be clearly understood that the basic predictions were plotted directly on the B.Z.C., whereas the predictions plotted on the 3-zone charts had been averaged over the range of longitudes 30° E to 30° W and combined with another range of longitudes from 150° E through 180° to 150° W. The error introduced in this averaging process may be illustrated in a practical use of the charts by plotting in Fig. 2 the diurnal curves for the circuit from

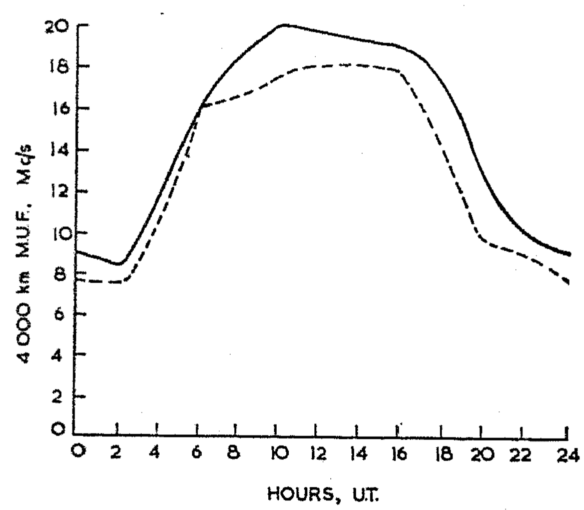


Fig. 2.—Predicted m.u.f. for London–Singapore circuit, September, 1953.
—— British zone prediction.
---- Zone 1 prediction.

London to Singapore for September, 1953. This graph refers to the London end of the circuit, which controlled throughout the day. The upper curve has been read from the B.Z.C. and the lower from the 3-zone chart. From this plot it is clear that a user who has a frequency of 17 Mc/s allocated to this circuit would read the predicted time of use from the latter chart as 0900–1630 hours U.T., whereas the basic prediction (also given on the B.Z.C.) showed this frequency to be satisfactory for 3½ hours longer (i.e. from 0700 to 1800 hours U.T.). Alternatively, if the user had a frequency of 19 Mc/s allocated to the circuit the 3-zone charts would show this to be too high, but the true prediction (as given on the B.Z.C.) gave the hours of operation on this frequency as 0900–1600 hours U.T.

It is axiomatic that, in the last case, a user who only wishes to communicate from 0900 to 1630 hours U.T. on 17 Mc/s or who has not a suitable frequency allocated could not have benefited in this particular case. However, communication conditions are continually changing, and it is quite certain that over a period every user would benefit from the more accurate presentation of the predictions. There have been so many cases reported in which frequencies higher than those read from the predictions on the 3-zone charts have been used that it seemed highly probable that, if comparisons were made with the actual predictions (on a B.Z.C.), a demonstrable improvement would result.

(4) EXPERIMENTAL RESULTS

Circuit data obtained by several users of D.S.I.R. predictions were compared with the two forms of chart (3-zone and B.Z.C.) over a period of six or seven months. The results obtained by three typical users are given in Table 1, from which it can be seen

Table 1

User No.	Comparisons	Standard deviations	
		3-zone method	B.Z.C. method
		%	%
1	Error in predicted frequency at median fade-out time	11·3	5·2
2	Error in predicted frequency at median fade-out time	18·3	9·2
3	(*a*) Error in predicted frequency at median fade-out time	19·2	12·7
	(*b*) Error in predicted fade-out time	2 hours	1·3 hours

that the standard deviation between the actual and predicted circuit performance was about twice as great when the predictions were based on the 3-zone charts as when they were based on the British Zone Chart.

(5) CONCLUSIONS

The general conclusion of the user organizations was that, over the area to which the chart applied, a worth-while improvement had been achieved.

The above comparison was made between the British Zone Chart and the corresponding part of the international 3-zone chart system. From June, 1955, the latter system, which has been adopted for world-wide forecasting by the D.S.I.R. for the past nine years, will no longer be used. In its place a system of 12 charts, each depicting the ionospheric conditions which govern the propagation of radio waves at each even hour Universal Time, will be introduced. The new charts should therefore provide a more accurate presentation of the forecasts at the hours specified.

The B.Z.C. could not replace that world system of 12 Universal Time charts owing to the limited area to which it applies. Over that area, however, it would enable a greater accuracy to be achieved by filling the gaps between the two-hourly Universal Time charts. It also has the advantages of being direct reading and of having a much larger scale.

(6) ACKNOWLEDGMENTS

The work was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

621.396.2 : 621.396.822 : 621.396.11

Paper No. 1708 R
Oct. 1954

A STUDY OF COMMERCIAL TIME LOST ON TRANSATLANTIC RADIO CIRCUITS DUE TO DISTURBED IONOSPHERIC CONDITIONS

By J. K. S. JOWETT, B.Sc.(Eng.), and G. O. EVANS, B.Sc., Associate Members.

(*The paper was first received* 2nd *March, and in final form* 24*th May*, 1954. *It was published in October*, 1954, *and was read before the* RADIO SECTION 9*th February*, 1955.)

SUMMARY

An analysis has been made of the time lost on certain radio circuits incoming to the United Kingdom from Montreal and New York during the period 1942–1952. The results have been examined in relation to the variations in solar and magnetic activity with a view to estimating the degree of interruption likely to be experienced on transatlantic circuits during the next few years. It is concluded that the maximum values of lost time occur during the winter months as the minimum of the sunspot cycle is approached, and that these high values are associated with periods of high magnetic activity arising from the occurrence of M-region storms. It is considered that the adverse propagation conditions encountered since the winter of 1950–51 will be slightly less severe during the winter of 1954–55 and that by the winter of 1955–56 there should be a marked reduction in the amount of commercial time lost on transatlantic circuits.

(1) INTRODUCTION

It has been known since the introduction of short-wave transmission for commercial communications that high-frequency radio-telephone and radio-telegraph services between the United Kingdom and North America are at certain times particularly subject to interruptions due to disturbed ionospheric conditions. These conditions are generally, but not invariably, associated with periods of high magnetic activity, when the values of the different components of the earth's magnetic field show considerable variations from their normal "quiet day" values. It is recognized that both high magnetic activity and adverse radio propagation conditions are symptoms of disturbed conditions in the ionosphere. The relation between these two effects is nevertheless of considerable interest, more particularly because radio engineers have consistently found it easier to relate radio propagation conditions to the degree of magnetic activity than to some primary characteristic of the ionosphere. A study of this correlation is further justified because of the help which existing records of magnetic activity, extending over a period of more than 50 years, might give in indicating future trends.

During sunspot-maximum conditions, periods of high magnetic activity are frequently associated with the occurrence of large sunspot groups on the sun's visible hemisphere; but during sunspot minima there are more general indications of other types of magnetic storms, believed to be associated with the so-called M-regions[1] on the sun. These storms tend to recur about every 27 days, and it is often possible to distinguish two or more series of storms within this period. Certain published information[2] and previous unpublished analyses of the time lost on North Atlantic telephony circuits have suggested that the highest values of lost time occur, not at the peak of the sunspot cycle, but generally about two or three years later. The present cycle, which is now approaching a minimum, had its maximum in mid-1947, but solar activity remained at very high near-maximum levels throughout the years 1947, 1948 and 1949. It might therefore have been expected that the highest values of lost time would occur at some time in the period 1949–1951.

It was not until September, 1950, however, that the lost time on circuits incoming from the United States and Canada showed any significant increase; at that time conditions on these circuits suddenly deteriorated, and the high values of lost time then recorded have been a feature of the conditions generally prevailing on them until the end of 1952. This increase in lost time coincided with a sharp decrease of solar activity and the onset of M-region magnetic storms recurring at intervals of 27 days. Reference to this deterioration in propagation conditions on North Atlantic circuits during the past two or three years has also been made in recent French[3] and Dutch[4] papers.

The high values of lost time on these circuits, sustained as they have been over the past three years, have had a serious effect upon the average traffic-handling capacity of individual circuits. The operational difficulties have, of course, been all the greater because accurate forecasts of general propagation conditions likely to be met 24 or 48 hours ahead are impossible, and short-term re-routing of traffic and readjustments of staffing arrangements can therefore rarely be made in advance of changing circuit conditions. While these difficulties emphasize the need for a reliable short-term storm-warning service, it is equally important at the present time to forecast the general conditions likely to prevail over the next few years.

As a first step it has been necessary to assemble information concerning the recent increases in lost time in a precise and useful form, and an examination has therefore been made of the records of certain North Atlantic circuits from 1942 onwards. By considering the facts thus obtained in relation to the prevailing magnetic and solar activity at different epochs of the sunspot cycle, an attempt has been made to estimate the future trend of lost time on these circuits as the minimum of the present sunspot cycle is approached.

(2) SOURCES OF DATA

(2.1) Circuit Data

Studies have been made (in the direction incoming to London) of three circuits—the New York–London and Montreal–London radio-telephone circuits and the Montreal–London radio-telegraph circuit. Records of lost time due to ionospheric disturbances on the two radio-telephone circuits, both of which employ the single-side-band method of transmission, have been obtained from logs kept by the London terminal station. In analysing these records, circuit time has been considered to be lost whenever the circuit was graded as unsuitable for connection to the inland telephone system. The New York circuit chosen for analysis used a high-power transmitter and the highly directional m.u.s.a. (multiple unit steerable array) receiving system;[5,6] this circuit has a normal 22-hour schedule of 1000–0800 hours G.M.T. daily. The Montreal circuit has a normal 17-hour schedule of 1200–0500 hours G.M.T. in summer and 1300–0600 hours G.M.T. in winter.

Logs of the performance of the Montreal–London radio-telegraph circuit, which is scheduled to operate for 24 hours

Mr. Jowett and Mr. Evans are in the Post Office Engineering Department, Radio Planning and Provision Branch.

daily except at week-ends, are kept by the receiving station at Somerton. The time lost on this circuit has been assessed by considering the periods during which the signals were graded as unreadable.

(2.2) Solar and Magnetic Data

Solar activity has been expressed by means of the 12-month running-average value of the Zürich monthly-mean sunspot number, R.

Magnetic activity, or disturbance of the earth's magnetic field, has been expressed by using the character-figures K[7] and C recorded by the Abinger magnetic observatory. K is an internationally agreed 3-hourly index giving the degree of magnetic activity over the periods ending at 0300, 0600, 0900 . . . 2400 hours G.M.T., and is obtained by taking the greatest of the three individual K-values indicating respectively the departures from quiet-day values of the horizontal and vertical components of the earth's magnetic field and the magnetic declination. C, which is referred to as the figure of the day, gives a measure of the degree of magnetic activity during the whole day and is proportional to the daily sum of the individual K values.

An examination has been made of K for the Lerwick (Shetland) and Cheltenham (Maryland, U.S.A.) magnetic observatories, to see whether either of them shows more evidence of magnetic activity (as expressed by a greater proportion of high K-values). The results of the examination have shown that the simultaneous values observed at each of the three observatories followed closely similar trends. Reference is also made by Bartels[7] to the high degree of correlation existing between values of K measured at different observatories.

(3) THE VARIATION OF MONTHLY LOST TIME ON THREE CIRCUITS WITH PARTICULAR REFERENCE TO SOLAR AND MAGNETIC ACTIVITY

(3.1) Presentation of Data

Figs. 1 and 2 show the variations in the monthly-average lost time on the New York–London and Montreal–London radio-telephone circuits and the Montreal–London radio-telegraph circuit over differing periods between 1942 and 1952.*

These Figures also show the variation of the 12-month running-mean Zürich sunspot number and the monthly-mean and yearly-mean values of C.

(3.2) Discussion of Results

(3.2.1) New York–London Radio-Telephone Circuit.

The results of the analysis for the New York–London radio-telephone circuit show that during the period analysed the lost time has its maximum value around the minimum of the sunspot cycle, and that only during certain months of the years of sunspot maximum (1947, 1948 and 1949) did the lost time fall to almost negligibly low values. The graphs also show peaks of lost time at the equinoxes, corresponding closely with marked increases in the degree of magnetic activity. This equinoctial increase in the degree of magnetic activity is a well-established phenomenon.[8] The equinoctial peaks of lost time are most clearly seen at sunspot maximum when, in fact, only the equinox periods cause any material difficulties in operation; during summer and winter in sunspot-maximum years the monthly lost time falls to less than 1% of the scheduled time.

(3.2.2) Montreal–London Radio-Telephone Circuit.

The general trend of lost time on the Montreal–London circuit shown in Fig. 1 is similar to that of the New York–London circuit although the values are considerably higher. This is to be expected, since the Montreal–London great circle penetrates more deeply into the North Auroral zone and consequently the signals are subject to greater absorption, especially during disturbed ionospheric conditions.

(3.2.3) Montreal–London Radio-Telegraph Circuit.

Fig. 2 shows the lost time on the Montreal–London radio telegraph circuit from 1944 onwards. In this case the lost time has been expressed as the monthly-mean time during which the circuit was graded as "signals unreadable or unheard" at the Somerton receiving station. Here again it is seen that the general trend of the lost time is similar to that on the New York–London radio-telephone circuit. There are, however, certain significant differences between the Montreal–London telephony and telegraphy circuits and these are discussed in Section 3.2.4.

(3.2.4) Comparison of the Montreal–London Radio-Telephone and Radio-Telegraph Circuits.

In Fig. 2 the percentages of time lost on both the Montreal–London radio-telephone and radio-telegraph circuits are compared. Since different standards have of necessity been used in assessing when time is lost, it is unwise to press the comparison too far, but it is of special interest to note the differing extents to which the two circuits follow the degrees of disturbance of the earth's magnetic field. It is seen that during all summers, and also during winter sunspot maximum, the lost time on both telephone and telegraph circuits follows the variations of magnetic activity very closely. During winter sunspot minimum, however, the telegraph-circuit lost time follows the magnetic variations much more closely than that of the telephone circuit. Examples of this trend can be seen during the winters of 1950–51 and 1951–52. Thus, considering the period from September, 1950, to May, 1951, the curve of the monthly-mean value of C shows two well-defined peaks in October and April, with a trough in which minimum values occurred in December and January. The curve of telegraph-circuit lost time shows similar variations, with one peak extending over October and November and a second peak in April; the minimum values occurred in December and January. For the telephone circuit the values remained high throughout the period October–April and did not show the pronounced reduction in December and January. A similar effect is seen during the winter of 1951–52.

This difference in the behaviour of the telephone and telegraph circuits during periods of high magnetic activity and winter sunspot minimum may, it is suggested, be due to differences in the characteristics of the aerials used for the two services. It is during sunspot minimum and particularly at night in winter that by far the greatest amount of lost time occurs; at these times the maximum usable frequencies for these circuits fall to low values during winter nights and there is a risk of signals penetrating the F layer of the ionosphere if they are radiated at too high an angle with the ground.

Fig. 3 shows the values of wave-arrival angle, for a frequency of 5 Mc/s, measured on the m.u.s.a. receiver at Cooling during each January for the period 1943–46, which includes sunspot minimum in 1944. Although this Figure refers to the New York–London circuit the results are thought to be generally applicable also to the Montreal–London circuits included in the study. It is seen that during 1943 and 1944 the average value of the wave-arrival angle was less than 15°, and during the pre-dawn period in 1944 much lower values were observed. A rise in angles can be noted in 1945, and by 1946 a sharp recovery to angles of about 20° occurred. This suggests that the aerials used on the transatlantic circuits should be designed to give their

* The paper was originally prepared early in 1953, but data for the whole of that year have since been added to Figs. 1 and 2. It will be seen that where these later results afford new information they support the conclusions arrived at in the paper.

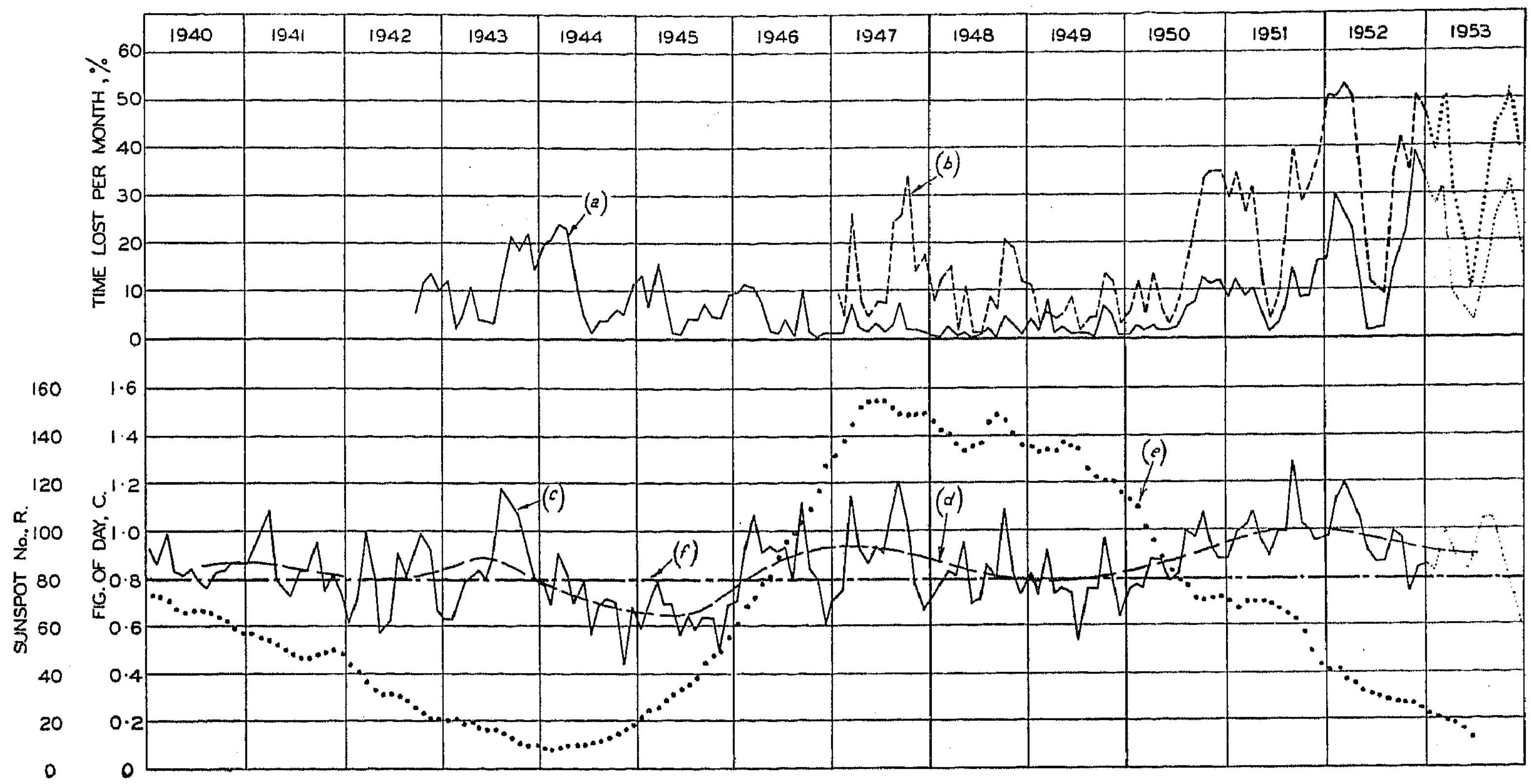


Fig. 1.—Percentage of scheduled time on New York–London (high-power system using m.u.s.a. reception) and Montreal–London telephony circuits (incoming direction only), and variations in solar and magnetic activity.

(*a*) New York–London.
(*b*) Montreal–London.
(*c*) Abinger magnetic figure of the day, *C*.
(*d*) Yearly average of *C*.
(*e*) Zürich sunspot number, *R*.
(*f*) Arbitrary reference level.

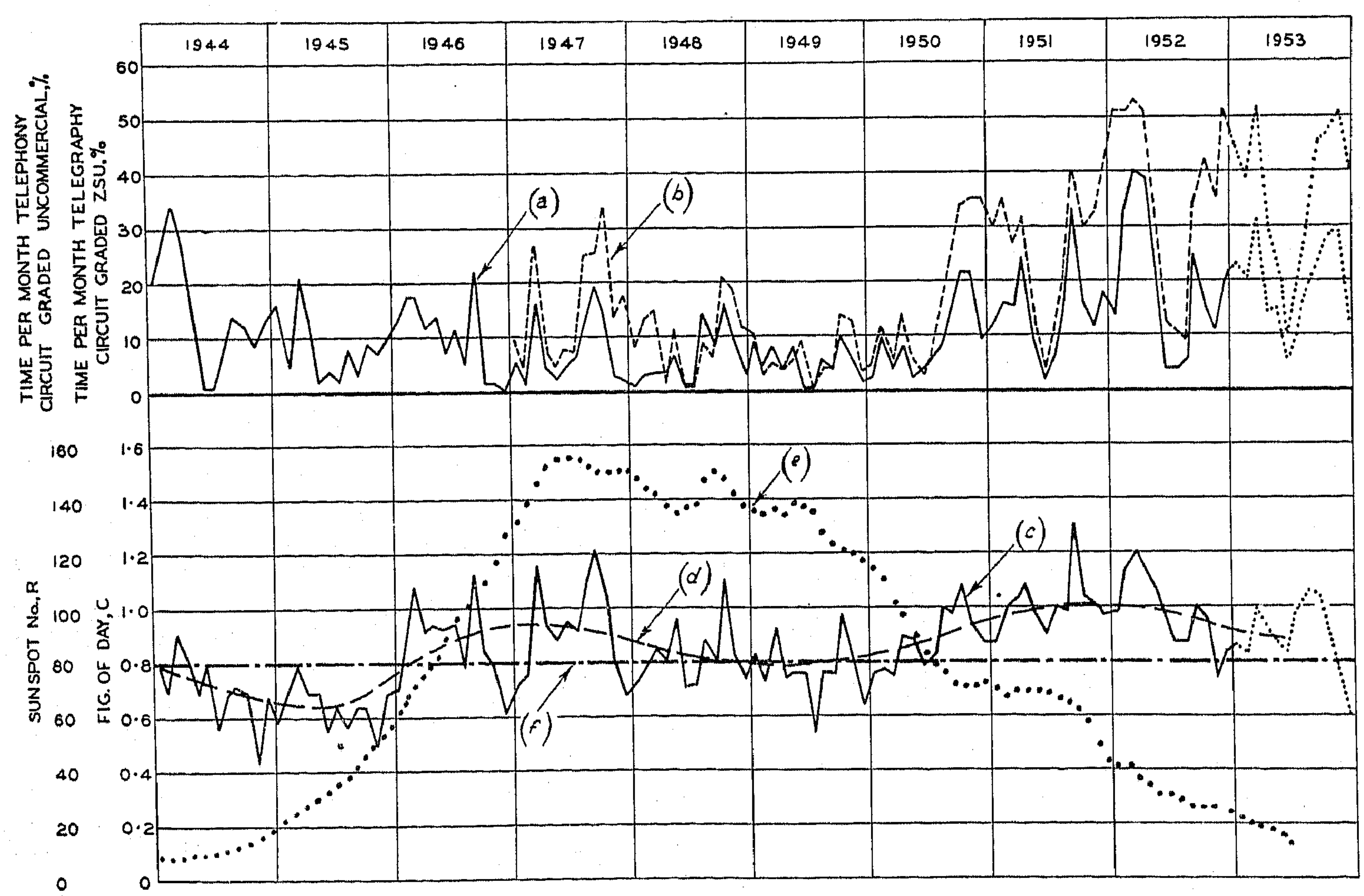


Fig. 2.—Percentage of scheduled time lost on Montreal–London telephony and telegraphy circuits (incoming direction only), and variations in solar and magnetic activity.

(*a*) Telegraphy circuit.
(*b*) Telephony circuit.
(*c*) Abinger magnetic figure of the day, *C*.
(*d*) Yearly average of *C*.
(*e*) Zürich sunspot number, *R*.
(*f*) Arbitrary reference level.

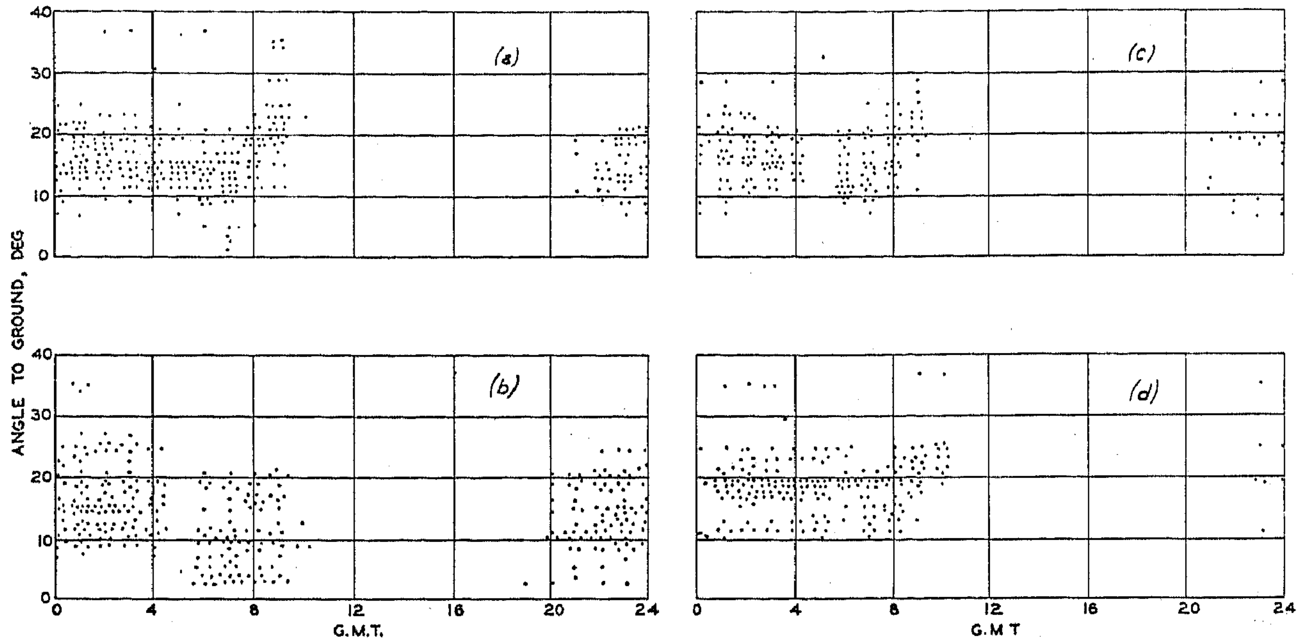


Fig. 3.—Wave arrival angles measured on m.u.s.a. receiver at Cooling for incoming 5·078-Mc/s transmission from Lawrenceville, New Jersey.

(*a*) January, 1943; (*b*) January, 1944; (*c*) January, 1945; (*d*) January, 1946.

maximum gain for frequencies of the order of 5 Mc/s at angles at least as low as 15° if sunspot-minimum conditions are to be adequately covered at night-time.

(4) THE RELATION OF TIME LOST DAILY DURING HOURS OF DARKNESS WITH MAGNETIC ACTIVITY AND SUNSPOT NUMBER

(4.1) Extent of Analysis

The greater part of the time lost on the transatlantic circuits occurs during the hours of darkness and between September and April. In order to examine further the relation which exists between the magnetic activity and lost time when daily, rather than monthly, mean values are considered, an analysis has been made of the time lost on the Montreal–London telephony circuit during the period 2100–0600 hours G.M.T. in relation to the midnight value of the magnetic index K, i.e. for the 3-hourly period 2100–2400 hours G.M.T. This particular period of the day has been selected because it is generally the most difficult from the point of view of transatlantic radio communication, and the analysis should in particular reveal the extent to which decreasing solar activity controls the amount of lost time on the circuit.

(4.2) Presentation of Results

The results of this daily analysis are shown in Fig. 4, which has been prepared in the following way. Taking first those days when one particular midnight value of K was measured, the total lost time over the period 2100–0500 hours G.M.T. during the difficult equinox months of September and October was deduced and expressed as a percentage of the total schedule time during this period. This procedure was repeated for all other integral values of K and the percentage thus obtained has been plotted against K. Separate curves have been drawn for each year between 1947 and 1952. Histograms, showing the number of occasions, N, on which each value of K occurred during the 2-month period, give a measure of the significance to be attached to each point on the curves. Under each of the Figures are shown also the mean values, over the 2-month periods concerned, of the following factors:

(*a*) 2-month running mean sunspot number.

(*b*) Calculated mean m.u.f. (maximum usable frequency) for the period 2300–0200 hours G.M.T.

(*c*) Mean value of magnetic index C (figure of the day).

The values of (*a*) and (*b*) show that while solar activity, and hence maximum usable frequencies, remained at very high values over the first three years (1947, 1948 and 1949) there was a sharp decline at the autumn equinox of 1950 and further progressive falls in the succeeding two years. Factor (*c*), the magnetic figure of the day, gives a useful indication of the extent to which the earth's magnetic field is disturbed, and is seen to have had exceptionally low values in 1949 and exceptionally high values in 1947 and 1951. This conclusion can to a large degree be drawn also from the histograms of K-values, but whilst the K-values relate only to a short period of the day (2100–2400 hours G.M.T.) the C figure is a measure of magnetic field variations throughout the whole of the day.

(4.3) Discussion of Results

The conditions occurring in 1947 being ignored for the present, a general survey of the Figures shows clearly the increase in lost time in the later years as the approach is made to sunspot-minimum conditions. Thus, for example, whilst the 1948 and 1952 magnetic-field conditions were not very different the lost time in 1952 was much greater, and considerable amounts of lost time occurred in 1952 at very low values of K, such as $K = 1$ and $K = 2$. There was, in fact, a steady worsening in radio conditions from 1948 to 1952 as solar activity declined, the relative degradation of circuit performance showing up markedly for the low values of K when conditions magnetically are relatively quiet.

The conclusion may be drawn that the night-time difficulties in the operation of the Montreal–London telephony circuit during 1950, 1951 and 1952 have arisen primarily from the large reduction in m.u.f. accompanied by very much smaller reduction in l.u.f. (lowest usable frequency) which has resulted in a low margin between m.u.f. and l.u.f. During disturbed ionospheric conditions there is a still further reduction in m.u.f. due to a fall in F-layer critical frequencies and an increase in l.u.f. resulting from increased absorption below the E layer.[9] Under these conditions it appears that ionospheric disturbances associated with even slight magnetic disturbances, corresponding to very low values of K, have been sufficient to make the circuit unworkable, or, at best, marginal in performance.

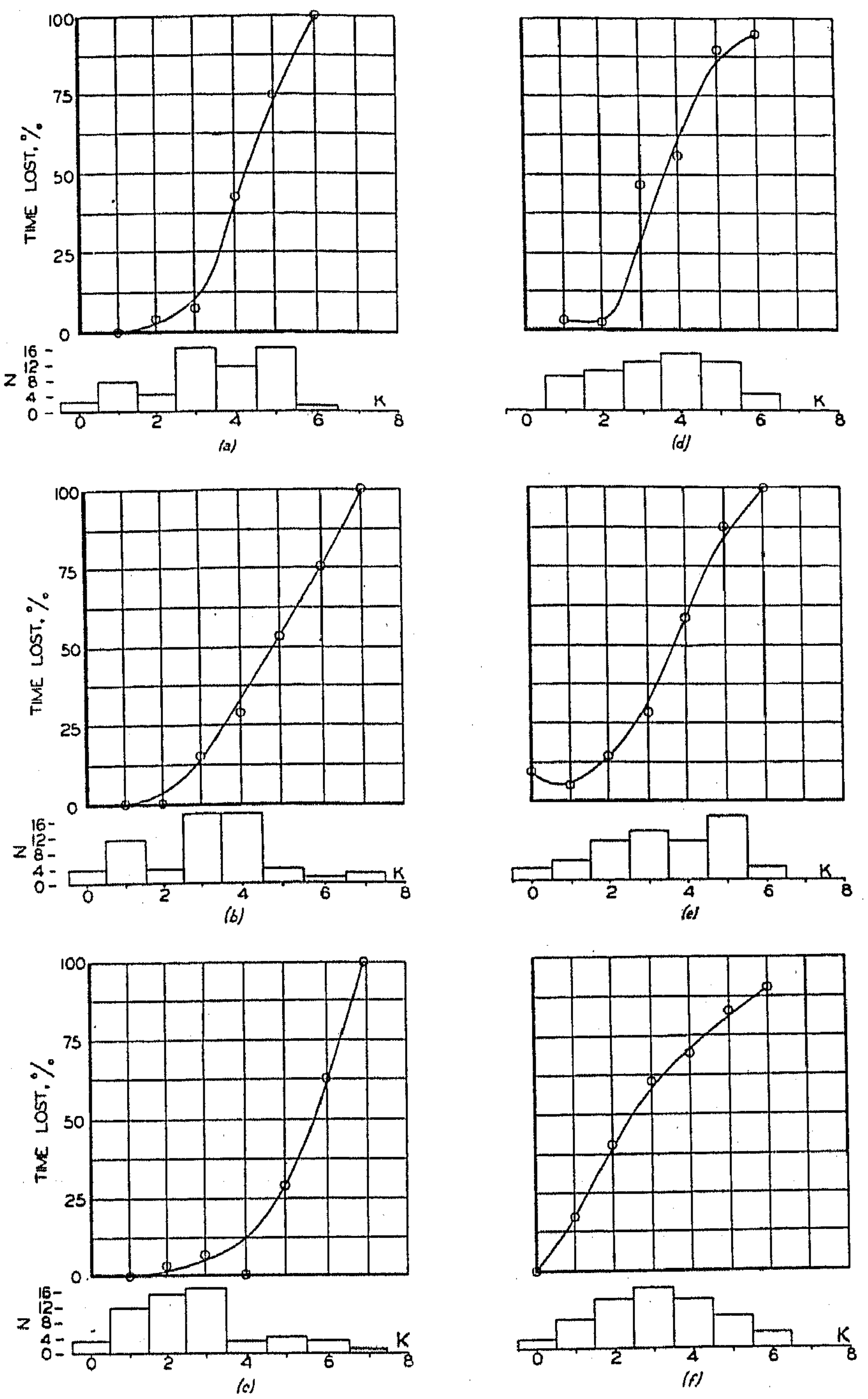


Fig. 4.—Montreal–London radio-telephone circuit. Relation between percentage of time lost between 2100 and 0600 hours G.M.T. daily and midnight value of magnetic index K during September and October for the years 1947–1952.

(*a*) 1947: m.u.f. = 10·55 Mc/s; C = 1·12; R = 150.
(*b*) 1948; m.u.f. = 12·35 Mc/s; C = 0·95; R = 149.
(*c*) 1949; m.u.f. = 12·10 Mc/s; C = 0·86; R = 122.
(*d*) 1950; m.u.f. = 8·75 Mc/s; C = 1·04; R = 71.
(*e*) 1951; m.u.f. = 8·70 Mc/s; C = 1·16; R = 61
(*f*) 1952; m.u.f. = 8·30 Mc/s; C = 0·98; R = 26.

When the year 1947 is considered, however, the conditions are seen to be very different. Lost time occurred only when K-values were moderately high, such lost time as there was coinciding almost entirely with values of 4 or 5. This may be attributed to the fact that near a sunspot maximum there is a large margin between the m.u.f. and l.u.f. It is of interest further to note that the lost time was especially high in the autumn of 1947, when difficult radio conditions frequently coincided with the visible passage of large sunspot groups. This type of storm, which does not generally seem to have a clear 27-day recurrence tendency, is believed to create greater ionospheric disturbances than the recurrent type, associated with so-called M-regions on the sun.[9]

It has been found that, on occasions, the lost time on the transatlantic circuits bears apparently little relation to the specific value of K prevailing at the time, although, as already shown, the correlation between monthly values of lost time and of C is high. Particularly is this true at the end of a severe magnetic storm lasting for several days; on such occasions it is not unusual for difficult radio-circuit conditions to continue for many hours, and sometimes even some days, after K has resumed very low values, corresponding to quiet magnetic conditions.

Whilst Fig. 4 shows several other interesting features, further evidence is required before more conclusions can be drawn; in particular, it would be of value to examine conditions during winter, and if opportunity permits this will be carried out at a later date.

(5) SOME RELEVANT FEATURES OF RECURRING MAGNETIC STORMS

Although the deterioration in the performance of the transatlantic radio circuits is of the same character whatever the type of long-term magnetic disturbance that accompanies it, the disturbances themselves fall into two main classes: (*a*) those associated with large sunspot groups, and (*b*) 27-day-recurring or M-region disturbances which have their origin in an active area on the sun's surface but which do not appear to be directly associated with sunspots.

During sunspot maxima the majority of the magnetic disturbances were associated with large sunspot groups and in particular with intense solar flares, which caused Dellinger fades followed by magnetic disturbances from 18 to 36 hours later. At sunspot maxima there is practically no evidence of disturbances clearly recurring after intervals of about 27 days.

As a sunspot minimum is approached the 27-day recurrence tendency becomes very pronounced, and during the present cycle the onset of the 27-day-recurrence storms coincided with the rapid decrease of sunspot number during 1950. This can be seen in Fig. 5, which shows the occurrence of magnetic-disturbance days from 1940 to the end of 1952 on a scale 27 days wide. Any recurrence tendency is shown as a series of vertical blocks, and it will be seen that for two or three years prior to July, 1950, there was little evidence of such recurrence tendency. The strength of the 27-day recurrence tendency has been examined by Shapley,[10] who concludes, from an analysis of data covering the years 1890–1944, that the 27-day recurrence tendency is most marked during years of decreasing solar activity and around sunspot minimum.

From Fig. 6, which shows the magnetic-disturbance days during 1952 together with the occurrence of large sunspots and solar flares, the recurrence tendency of the disturbance days can be clearly seen. It will also be seen that a marked recurrence tendency exists for the days during which the lost time on the New York–London circuit exceeded 15% of scheduled time.

(6) THE FUTURE TREND OF LOST TIME ON NORTH-ATLANTIC CIRCUITS

It may be assumed that sunspot-minimum conditions will persist at least until the spring of 1956, with some possibility of the actual minimum occurring during 1955. This would imply that there is a likelihood of the lost time remaining at the present high values, or even increasing, during the next two or three winters; some support for this view is given by the lost-time figures for December, 1952, which are amongst the highest yet recorded, although the monthly-average value of magnetic activity was not unduly high.

Lost time, however, has been shown in earlier Sections of the paper to depend not only upon the sunspot number but also upon the degree of magnetic activity. It will be seen from Fig. 1 that the magnetic activity fell to its lowest value for the period 1940–1952 in 1945, the year following sunspot minimum, although 1943, the pre-minimum year, was one of high magnetic activity. In fact, as shown in Fig. 5, the 27-day or M-region

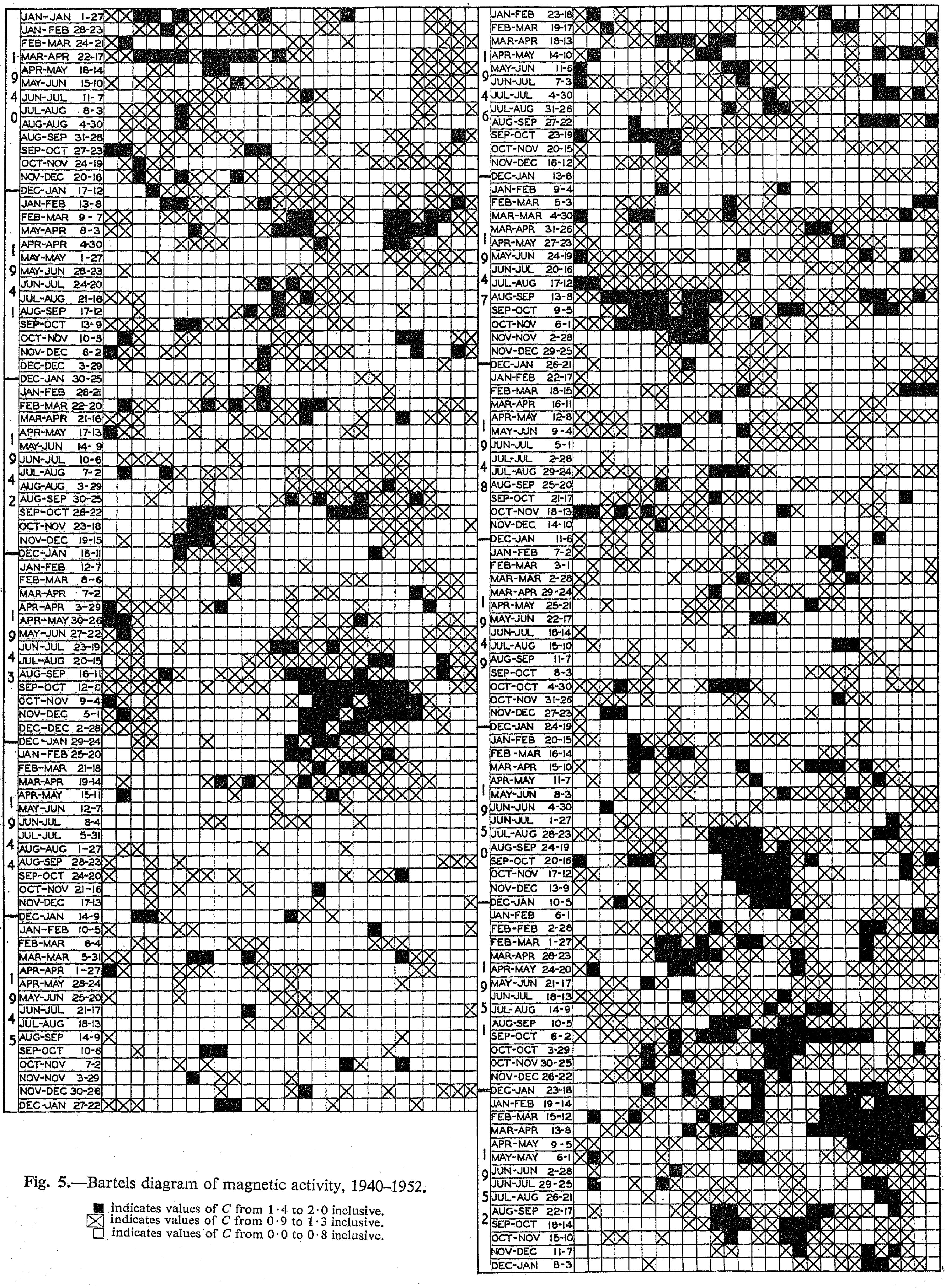


Fig. 5.—Bartels diagram of magnetic activity, 1940–1952.

■ indicates values of C from $1 \cdot 4$ to $2 \cdot 0$ inclusive.
☒ indicates values of C from $0 \cdot 9$ to $1 \cdot 3$ inclusive.
☐ indicates values of C from $0 \cdot 0$ to $0 \cdot 8$ inclusive.

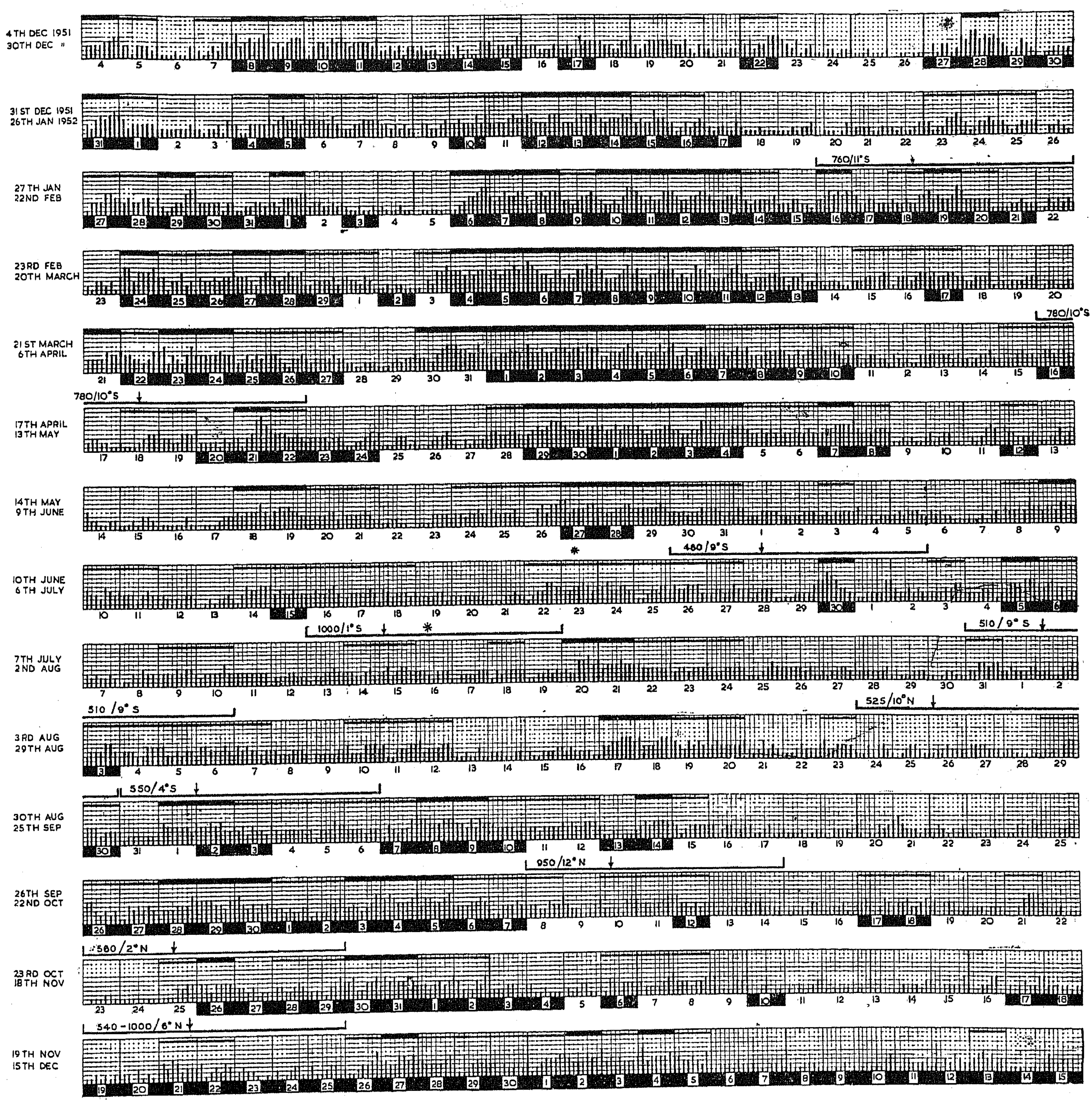


Fig. 6.—Sunspot and magnetic data, 1952.

indicates sunspots of area greater than 500 × 10^{-6} of visible hemisphere. Area and latitude are shown to left of arrow, which indicates central meridian passage.
indicates solar flare.

indicates values of C from 0·9 to 1·3 inclusive.
indicates values of C from 1·4 to 2·0 inclusive.
⊥ indicates 3-hourly magnetic index K: One small division = 1 unit.
indicates days on which lost time on high-power system using m.u.s.a. reception was greater than 15% of scheduled time.

storms, which account for the greater part of magnetic activity as sunspot minimum is approached, are no longer evident after the spring equinox of 1944. As mentioned in Section 5, the 27-day recurrence tendency is most marked during years of decreasing solar activity and around sunspot minimum.

The tendency for the magnetic activity to fall to its lowest value at sunspot minimum is also shown in the results of a study by Chapman[8] of the variations of magnetic activity and sunspot number during the period 1837–1930. It is found that there is comparatively little variation in the minimum values of magnetic activity for each cycle, although considerable variations occur in the maximum values. An examination of magnetic data for the period August, 1932–August, 1933 (the International Polar Year),[11] which was also a period near to sunspot minimum, shows that the degree of magnetic activity during this period was of the same low order as that during 1944 and 1945. It was only in 1842 that a high level of magnetic activity was reached at a sunspot minimum.

It therefore appears that, although it is not possible to forecast with certainty the degree of magnetic activity during the next few years, it is not unreasonable to assume that it will fall to a level similar to that recorded after previous sunspot minima. This period of low magnetic activity may, however, be preceded by a period of high activity in the pre-minimum year, and it will also be a period of low m.u.f.'s with a consequent small margin between the m.u.f. and l.u.f. It has already been mentioned in Section 4.3 that when there is a small margin only between the m.u.f. and l.u.f. the ionospheric disturbances associated with even slight magnetic disturbances are sufficient to make the transatlantic circuits unworkable, or, at best, marginal in performance, more particularly at night. The conclusions to be drawn from all these considerations are that the high values of lost time which have been characteristic of the winters of 1951–52 and 1952–53 will persist during the winter of 1953–54* in spite of a possible decrease in magnetic activity. It is impossible to say whether magnetic activity, as measured by the monthly-mean values of *C*, will have fallen sufficiently by the winter of 1954–55 to allow conditions to improve in spite of the low values of m.u.f. It might, however, be expected that by the winter of 1955–56 the minimum of the present sunspot cycle will have been passed. On the assumption that this is so, the anticipated low values of magnetic activity, together with the proportionally greater increase of m.u.f. as compared with l.u.f., should result in materially improved conditions.

(7) CONCLUSIONS

An analysis has been made of the percentage of scheduled time lost per month on radio-telephone and radio-telegraph circuits incoming at London from Montreal, and a radio-telephone circuit incoming at London from New York, for various periods from 1942 onwards. From a study of solar and magnetic data it is concluded that the present high level of lost time on these circuits is associated with a high degree of magnetic activity arising from the occurrence of M-region storms. These high values have persisted throughout the winter of 1952–53 and are expected to continue through the winter of 1953–54, but some slow improvement in conditions may be expected during 1954. No great improvement in conditions can, however, be expected until the commencement of the next sunspot cycle and, although this cannot yet be forecast with any accuracy, it is not expected to occur before 1955.

* Experience during the winter of 1953–54 confirmed these forecasts in that, although magnetic activity declined, the lost time on circuits between the United Kingdom and North America proved to be as high as in the two preceding winters.

(8) ACKNOWLEDGMENTS

The authors wish to thank the Engineer-in-Chief of the Post Office for permission to publish the paper, and to express their appreciation of the assistance given by several colleagues, particularly Mr. F. Morris, in the analysis and preparation of the available material. They also wish to acknowledge the assistance given by Cable and Wireless Ltd. in providing analysed observations for the Montreal–London radio-telegraph circuit.

(9) REFERENCES

(1) BARTELS, J.: "Terrestrial-Magnetic Activity and its Relations to Solar Phenomena," *Terrestrial Magnetism and Atmospheric Electricity*, 1932, **37**, p. 1.

(2) ANDERSON, C. N.: "Notes on the Time Relation between Solar Emission and Terrestrial Disturbances," *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 503.

(3) MAIRE, J.: "Note sur les relations entre l'agitation du champ magnétique terrestre et l'agitation ionosphérique observée dans les radio-communications transcontinentales par ondes 'décamétrique,'" *Annales de Géophysique*, 1949, **5**, p. 216.

(4) VAN SABBEN, D.: "Relationship between Radio Propagation Disturbances, Geomagnetic Activity and Solar Noise," *Journal of Atmospheric and Terrestrial Physics*, 1953, **3**, p. 194.

(5) GILL, A. J.: Wireless Section, Chairman's Address, *Journal I.E.E.*, 1939, **84**, p. 248.

(6) BOOTH, C. F.: "The Receiving System at Cooling Radio Station," *Post Office Electrical Engineers' Journal*, 1949, **42**, p. 84.

(7) BARTELS, J., HECK, N. H., and JOHNSTON, H. F.: "The Three-Hour-Range Index Measuring Geomagnetic Activity," *Terrestrial Magnetism and Atmospheric Electricity*, 1939, **44**, p. 411.

(8) CHAPMAN, S.: "The Earth's Magnetism" (Methuen's Monographs on Physical Subjects), p. 105.

(9) MINNIS, C. M.: "Ionospheric Storms and Radio Circuit Disturbances," *Wireless Engineer*, 1951, **28**, p. 43.

(10) SHAPLEY, A. H.: "The Recurrent Tendency and Forecasting of Magnetic Activity," *Transactions of the American Geophysical Union*, 1947, **28**, p. 715.

(11) "Geomagnetic Index K. International Polar Year, August, 1932–August 1933" (International Union of Geodesy and Geophysics. Bulletin 12d).

[The discussion on the above paper will be found on page 522.]

621.396.2/.3 621.396.11

Paper No. 1787 R
Jan. 1955

PERFORMANCE CHARACTERISTICS OF HIGH-FREQUENCY RADIOTELEGRAPH CIRCUITS

By A. M. HUMBY, Associate Member, C. M. MINNIS, M.Sc., F.Inst.P., and R. J. HITCHCOCK, B.A., Associate Member.

(The paper was first received 22nd October, and in revised form 3rd December, 1954. It was published in January, 1955, and was read before the RADIO SECTION *9th February, 1955.)*

SUMMARY

Operational data from several important long-distance radiotelegraph circuits have been analysed and the results discussed in relation to variations in solar and magnetic activity. It is shown that the lowest operational efficiency on radio circuits need not necessarily occur at the same time as the lowest level of solar activity. For example, in 1952 the performance of many difficult radio circuits deteriorated considerably, coincident with a peak of magnetic activity, but subsequently recovered although solar activity continued to fall.

On certain circuits there is a pronounced seasonal asymmetry of performance in the two directions of the route; to confirm that this is due to propagation effects, data from different organizations operating on substantially the same radio path are compared—particular reference being made to the United Kingdom–South Africa and the United Kingdom–Australia circuits. Suggestions are put forward as to the possible reasons for this asymmetry with season, and consideration is given to effects such as ionospheric propagation, atmospheric noise and the path taken by the ray.

Mention is also made of other propagation phenomena, such as echo signals, maximum latitude of ray path and M reflections, which have a bearing on the operation of radiotelegraph circuits.

For the purposes of the paper all references to seasons relate to the northern hemisphere.

TERMINOLOGY

Certain terms used in the paper, which have no generally accepted meanings or which are used here in a special sense, are defined with a view to avoiding misunderstanding.

Circuit performance.—The efficiency of a circuit as measured, for example, by the number of commercial hours per day or days per month.

Long and short routes.—The major and minor arcs, respectively, of the great circle passing through two points on the surface of the earth.

Forward echo.—A signal, arriving after the wanted signal, which has encircled the globe one or more times. Time intervals between the wanted and echo signals will be a multiple of approximately one-seventh of a second.

Backward echo.—The signal traversing the long (or short) route, at a time when the operation of the circuit is by way of the short (or long) route.

Reciprocity.—The reciprocity theorem implies that if a transmitter T is matched into an aerial A_T, and a receiver R is matched into a different aerial A_R at a distant point, the power generated in the receiver input impedance when this arrangement is used is the same as in the alternative arrangement in which T and R are interchanged and matched to A_R and A_T, respectively. The term *non-reciprocity* can therefore be applied only to differences in circuit performance which are due to propagation effects, and cannot be used to describe differences which are due to the terminal equipment or its utilization.

Asymmetry in circuit performance.—A difference between the performance of a circuit in the incoming and outgoing directions. The difference may be due to one or more causes and does not necessarily imply non-reciprocity.

Long-term asymmetry.—Asymmetry which remains constant over a long period.

Seasonal asymmetry.—The component of the asymmetry which varies regularly with season.

M Reflections.—Multi-hop ionospheric propagation in which the intermediate reflections take place between two ionospheric layers and not between a layer and the ground.

Percentage time commercial.—The percentage time for which the signals were suitable for printing at high speed.

(1) INTRODUCTION

In the early days of radiocommunication the frequencies employed were in the low- and very-low-frequency bands, and the realization that the radiation was guided round the earth by means of the ionosphere was consequently of little practical value to the user of a telegraph circuit.

With the advent of high-frequency (h.f.) communication about 30 years ago it soon became apparent that on a given radio circuit the range of frequencies which could be used at any particular time was determined by the ionosphere. Since there were no ionospheric forecasting services, the operation of existing and the planning of new circuits were largely matters of experience, notably in regard to the way in which the ionosphere, under both quiet and disturbed conditions, affected h.f. communications.

Whilst some of the results of experience can now be understood in terms of modern knowledge, it is worth recording certain aspects thereof which may not be generally known or for which no complete explanation exists.

It has been the practice over many years for operating personnel regularly to assess the merit of their circuits in one way or another, and such assessments, when analysed in conjunction with the prevailing solar, magnetic and other data, have been of the utmost value in the development of efficient radiotelegraph circuits.

The circuits referred to in the paper are illustrated in Fig. 1, and the sources of data are listed in Table 1.

In the years immediately following the initial operation, in 1927, of the high-speed radiotelegraph Empire beam circuits from this country to Australia, Canada, India and South Africa, certain performance characteristics emerged from the analysis of the merit assessments; for example,

(*a*) Effects associated with magnetic and solar activity.
(*b*) Asymmetry of performance, with season, of the two directions of a given radio path.

The subsequent increase in the number of high-frequency radio

Mr. Humby is at the Admiralty, and was formerly with Cable and Wireless, Ltd.
Mr. Minnis is at the Radio Research Station, Department of Scientific and Industrial Research.
Mr. Hitchcock is with Cable and Wireless, Ltd.

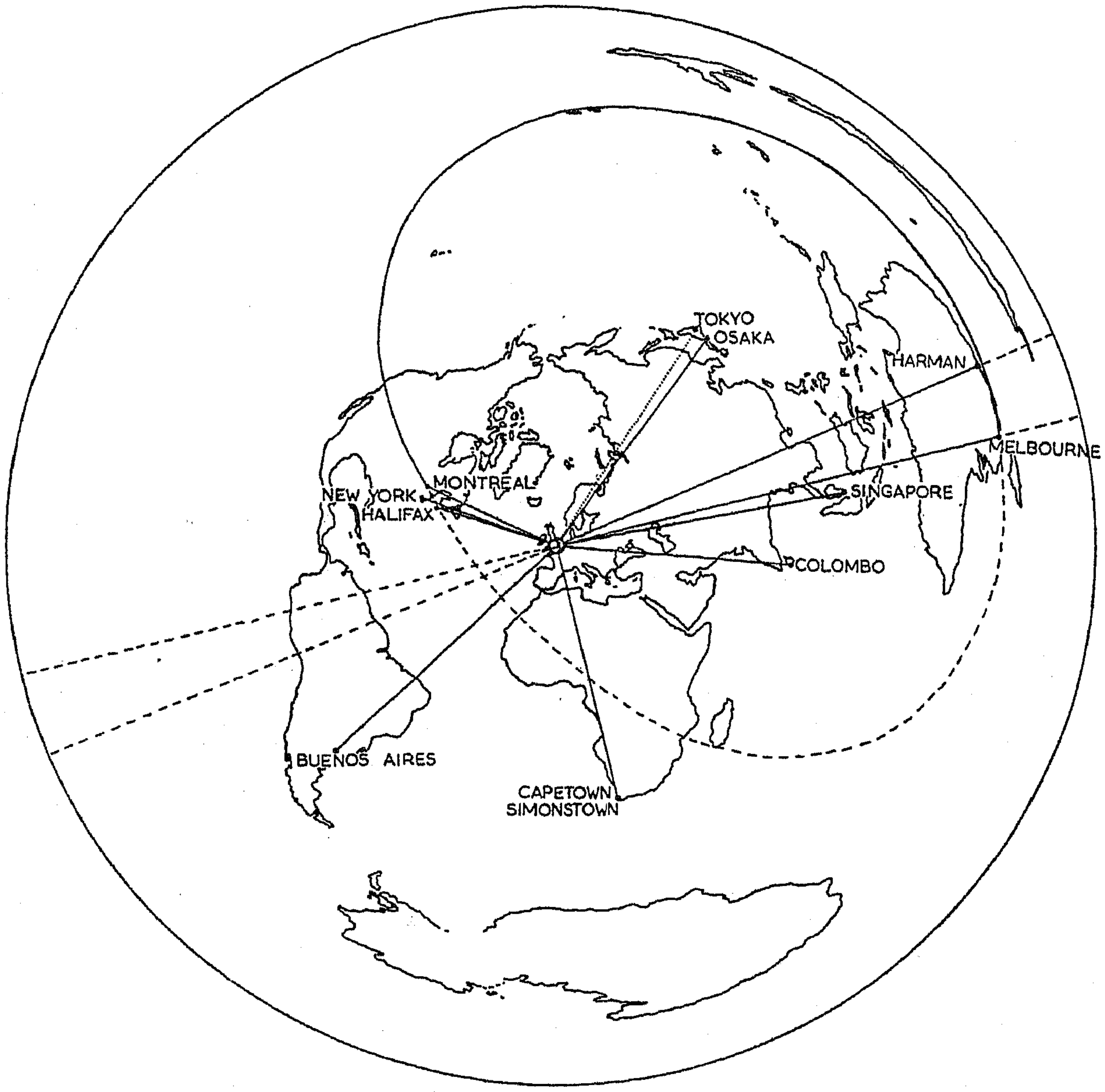


Fig. 1.—Great-circle routes of circuits in Table 1.

Table 1

Circuit	Source of data
London–Simonstown (Cape Town)	Admiralty
London–Singapore	Admiralty
London–Halifax (Canada) ..	Admiralty
London–Harman (Canberra) ..	Admiralty
London–Colombo	Admiralty and Air Ministry
London–New York	Cable and Wireless, Ltd.
London–Tokyo	Cable and Wireless, Ltd.
London–Melbourne	Cable and Wireless, Ltd.
London–Cape Town	Cable and Wireless, Ltd.
London–Buenos Aires	Cable and Wireless, Ltd.
London–Montreal	Canadian Overseas Telecommunications Corporation
Montreal–Melbourne	Canadian Overseas Telecommunications Corporation

circuits has enabled many more data to be assembled and, in certain instances, the performance of main-line circuits operated by different organizations over substantially similar radio paths to be studied.

For such circuits, with first-class equipment, the effects due to differences in sites, keying systems, continuity of skilled personnel, traffic loading, frequencies, interference, etc., are usually small in relation to those due to ionospheric effects. Such a case is illustrated in Fig. 2, in which the performance of the Admiralty and Air Ministry circuits from Colombo to London for the winter of 1952–53 are compared.

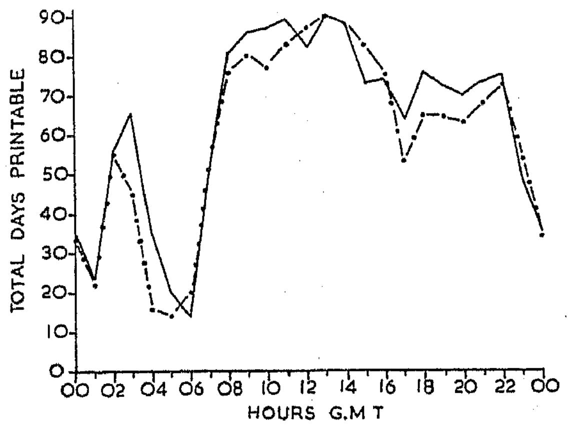


Fig. 2.—Comparison of performance over the same route of circuits operated by different organizations.

Colombo–London. November, December, January, 1952–53.
——— Admiralty.
·—·—· Air Ministry.

(2) EFFECTS ASSOCIATED WITH MAGNETIC AND SOLAR ACTIVITY

Of the four Empire beam circuits referred to in Section 1, the Canadian was found to be most susceptible to magnetic-storm effects. From 1928 to 1932 this circuit was operated on frequencies of approximately 9 and 18 Mc/s and was undoubtedly adversely affected by the absence of additional frequencies [Fig. 3, curve (*a*)]. The subsequent facilities for operation on any of the frequencies 4, 7, 9, 13 and 18 Mc/s, together with improvements such as spaced-aerial diversity reception and frequency-shift keying, resulted in a significant

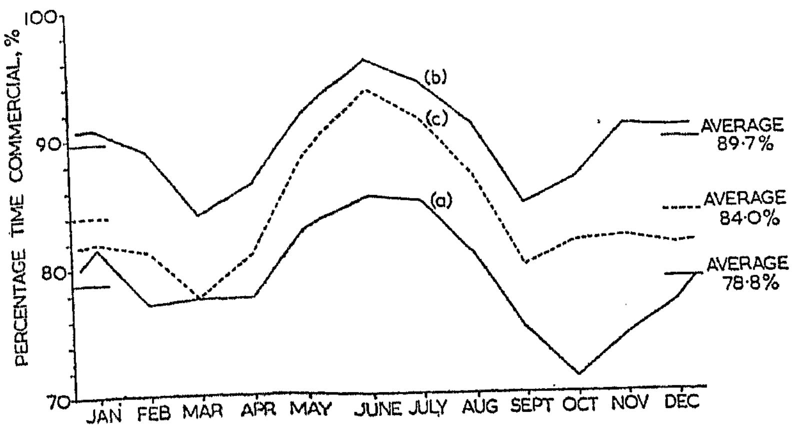


Fig. 3.—Seasonal change in performance of the London–Montreal circuit.

(a) Montreal–London, 1928–32 (inclusive); 9 and 18 Mc/s only.
(b) Montreal–London, 1933–52 (inclusive); 4, 7, 9, 13 and 18 Mc/s.
(c) London–Montreal, 1933–52 (inclusive); 4, 7, 9, 13 and 18 Mc/s.

improvement in performance [Fig. 3, curve (b)]; in curve (c) the performance in the opposite direction is compared for the same 20 years as that for curve (b). Best conditions are seen to occur in summer, contributory factors being the larger number of daylight hours and therefore greater stability of route, together with the improvement in transmitting- and receiving-aerial performances consequent upon the extended use of higher frequencies at this season.

The relatively poor performance during the vernal- and autumnal-equinox months is largely attributable to the high incidence of magnetic storms at these two seasons.

Whilst frequencies of 18 Mc/s and above are essential for daytime operation of the transatlantic route in sunspot-maximum years, the useful frequency range tends to be approximately 4–14 Mc/s in sunspot-minimum years; in fact, the 13–14 Mc/s band has proved extremely useful on this route at some period of the 24 hours at *all* phases of the sunspot cycle. This variation in the usage of a given frequency may, on some routes, change appreciably from month to month and year to year; compare, for example, the use on the Tokyo–London route of 9 Mc/s in January, 1934 and 1937 (see Fig. 4, curves A) and likewise of

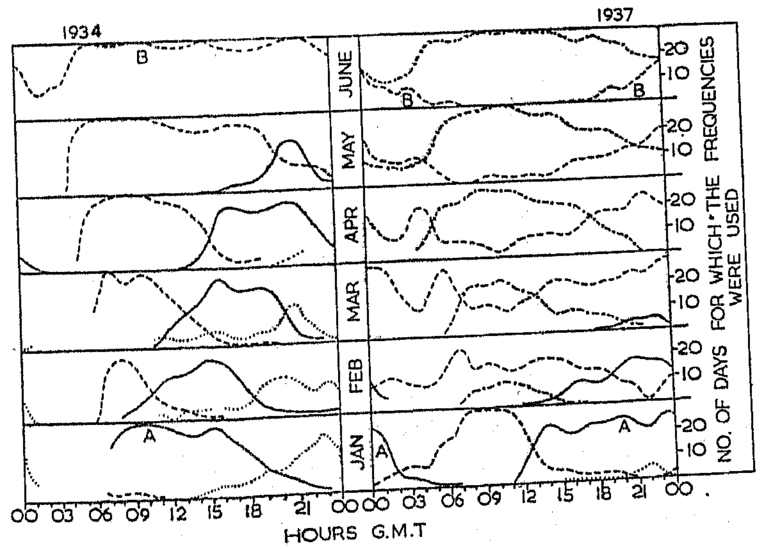


Fig. 4.—Frequency usage as a function of time of day and season for sunspot minimum (1934) and maximum (1937) for the Tokyo–London circuit.

.... 7 Mc/s.
——— 9 Mc/s.
------ 13 Mc/s.
—.—.— 18 Mc/s.

13 Mc/s in June, 1934 and 1937 (curves B). Within the three years, 1934–37, solar activity changed from an extremely low to a peak value; the years 1944–47 were associated with even greater changes in solar activity and therefore in frequency usage. Such changes illustrate the great difficulty, at least on an international basis, of producing satisfactory time schedules for non-simultaneous sharing of specific frequency assignments.

Owing to the restricted frequency range in sunspot-minimum years there may be little margin on certain routes for decrease of frequency under magnetic-storm conditions.[1] In consequence many circuits may suffer severe interruptions during years of low solar activity, e.g. 1943–44 (see Fig. 5).

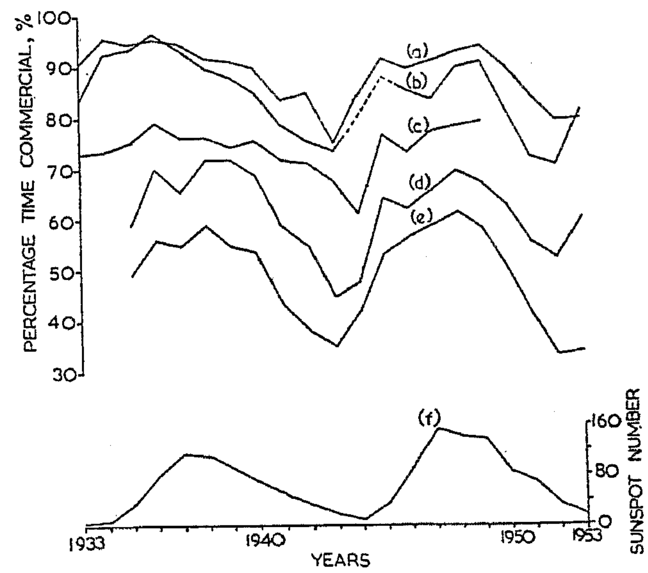


Fig. 5.—Circuit performance as a function of solar activity.

(a) Montreal–London, 1933–53 (inclusive).
(b) London–Montreal, 1933–53 (inclusive).
– – – – Amended curve, assuming that 4 Mc/s was available.
(c) Melbourne–London, 1933–49 (inclusive).
(d) Melbourne–Montreal, 1935–53 (inclusive).
(e) Montreal–Melbourne, 1935–53 (inclusive).
(f) Annual sunspot number.

In regard to curve (b) of Fig. 5, the frequency of 4 Mc/s was not available for use from London to Montreal in 1944 and 1945, whereas it was used successfully in the opposite direction [curve (a)] for 12·7% of time in 1944 and 10·2% of time in 1945.

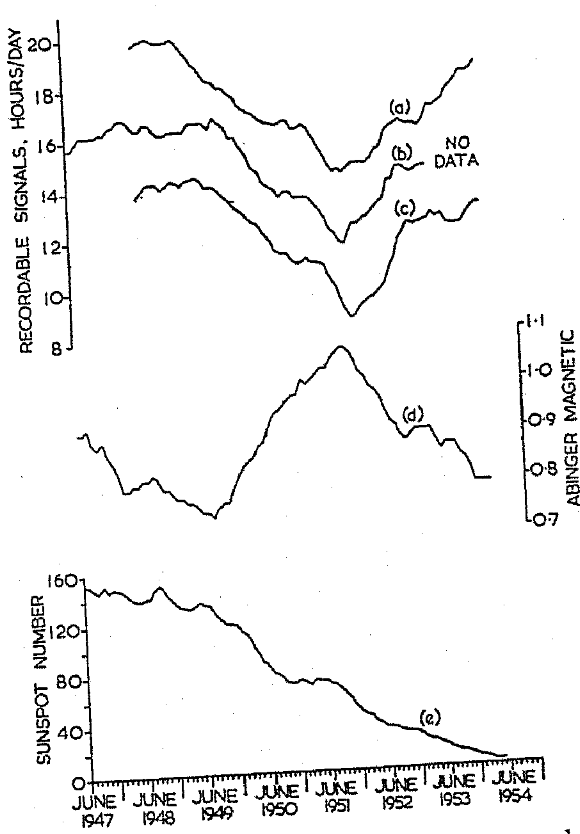


Fig. 6.—Relationship between circuit performance and magnetic and solar activity (12 months' running means).

(a) Halifax–London.
(b) Melbourne–Montreal.
(c) Harman–London.
(d) Geomagnetic activity at Abinger, C values.
(e) Solar activity.

Assuming similar usage in the direction London–Montreal, the performance in the two directions would have been substantially similar. Maximum interruption and minimum solar activity may not necessarily be coincident in time; for example, the performance of many difficult circuits had deteriorated appreciably by the middle of 1952 (see Fig. 6) simultaneously with a peak of magnetic activity, but recovered subsequently despite a continued decrease of solar activity.

(3) SITING OF RADIO RELAY STATIONS

The interruptions mentioned above refer only to the direct circuits concerned and can, of course, be materially reduced by the use of a suitably-sited relay station.[1]

(3.1) Effect of Atmospheric Noise

In attempting to avoid auroral zones in the selection of sites for relay stations, difficulties may arise owing to the higher noise levels associated with certain areas in temperate and equatorial latitudes.

In this connection, investigations in the summer of 1935 and the winter of 1935–36 into the comparative merits of Ascension Island and Sierra Leone for a relay station for the London–Montreal route disclosed that improved signal/noise ratios were obtained at Ascension Island, particularly during summer nights.[2]

(3.2) Effect of Sudden Ionospheric Disturbances

Difficulties may also be encountered on non-auroral paths owing to high absorption of signal during daylight over the path as a result of sudden ionospheric disturbances.[3] This effect is greatest during years of high solar activity since solar eruptions may give rise to prolonged periods of poor signal strength. Since this condition is essentially associated only with the sunlit hemisphere, facilities both for east- and west-about relay systems are advantageous owing to the complementary characteristics of the two directions.

Interruptions due to sudden ionospheric disturbances increase seriously as the frequency in use is decreased; in consequence, full advantage should be taken of the fact that on many routes, such as Singapore–Colombo, frequencies of the order of 20 Mc/s or above are suitable for a large proportion of the solar cycle owing to the proximity of such routes to the equatorial peak of critical frequencies.

In such an event, not only would the interruptions from sudden ionospheric disturbances be reduced to a minimum, but the circuits would be much less susceptible to interference owing to (*a*) the higher realizable signal gains and signal/interference ratios obtainable at these higher frequencies and (*b*) the fact that for many periods such frequencies would be in excess of those on which propagation could take place from sources in higher latitudes, e.g. Western Europe, to the receiving terminals in question.

(4) ECHO SIGNALS

Reference has already been made to the difficulties encountered in practice in regard to non-simultaneous sharing of individual frequency assignments, and although a limited degree of simultaneous sharing is practicable, notably at the lower end of the h.f. band, it has been known from the earliest days of regular h.f. communication that difficulties may arise elsewhere in the band owing to the presence of backward and forward echoes which may occur under certain propagation conditions.[4, 5]

(4.1) Forward Echo

Fig. 7 shows an example of forward-echo reception at Bridgwater, consisting of two consecutive dots A_1 transmitted from

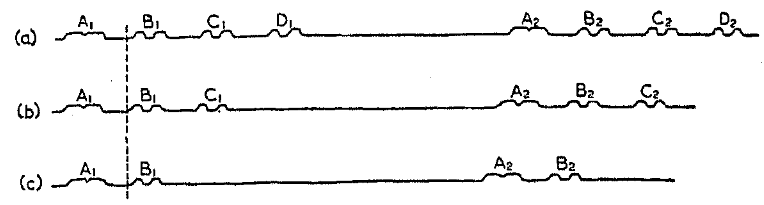


Fig. 7.—Forward echoes on the Cape Town–London circuit.

Transmission Cape Town, 18 660 kc/s.
Reception Bridgwater, Somerset, 31st October, 1938.
(*a*) Triple echo: 1718 hours G.M.T.
(*b*) Double echo: 1820 hours G.M.T.
(*c*) Single echo: 1910 hours G.M.T.
Transmitter quiescent during interval A_1, A_2.

Cape Town, together with a sequence of triple echoes (B_1, C_1, D_1) of the main transmission as a result of the rays encircling the globe three times. The number of echoes progressively decreased, reception being restored to normal after about three hours. A limited number of measurements between 1700 and 1900 hours G.M.T. on 31st October, and 1st and 2nd November, 1938, resulted in the following order of magnitudes of these echoes referred to the level of the first signal received:

First echo	−26 dB
Second echo	−39 dB
Third echo	−42 dB

The above effect may occur whenever the complete great-circle path appropriate to the transmitting and receiving terminals is mainly, or wholly, in twilight, and may pass unnoticed if multi-path effects are not excessive and if differential fading of the signals is sufficiently small to enable the receiver automatic-gain-control circuits to suppress the echoes.

(4.2) Backward Echo

In contrast to the above, backward echo may continue for several hours and produce severe telegraph distortion. Fig. 8

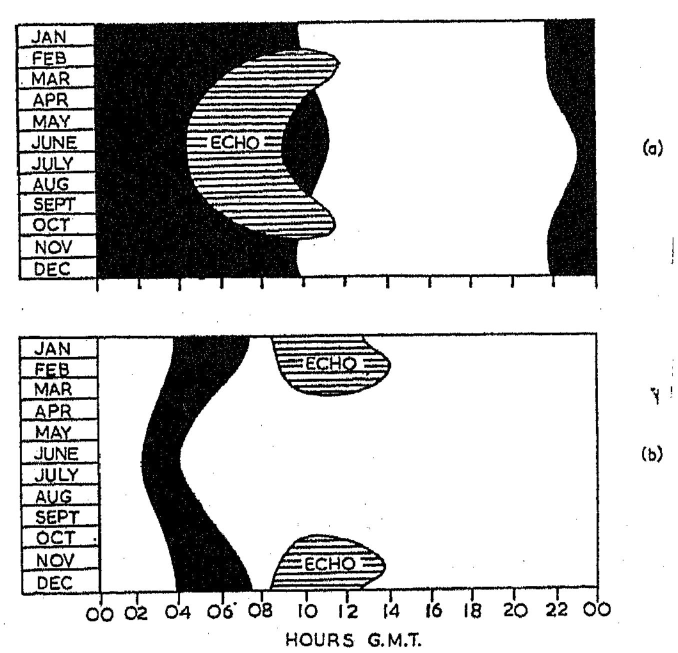


Fig. 8.—Effect of solar activity on reception conditions Java-Berlin, 8–20 Mc/s.

(*a*) 1927–29: Sunspot maximum.
(*b*) 1933–34: Sunspot minimum.
Unreliable periods.
Presence of backward echo.
(Extracted from C.C.I.R. Documents, Bucharest, 1937.)

compares the unreliable periods due to this and other effects for the Java–Berlin route for a maximum and minimum epoch of the solar cycle.

The Singapore–London route is very similar, and backward

echo may likewise be very prevalent when signals arriving over the short route are heavily absorbed owing to daylight, in contrast to the low absorption associated with the long route under its night conditions.

In the early days of h.f. radiotelegraphy, i.e. when it was normal practice to use slow-speed Morse or revert to it when high-speed signalling suffered from unacceptable distortion, echo was readily detectable by aural methods. In these days, however, systems employing frequency-shift keying, and channelling in terms of time and/or frequency, render diagnosis of echo difficult [see Fig. 11(*b*)], and the signal distortion therefrom may be attributed to other causes, such as faulty transmission or interference from unidentified sources, unless investigations are carried out by the use of special test signals and methods of reception such as moving film or undulator. Fig. 9(*a*) shows

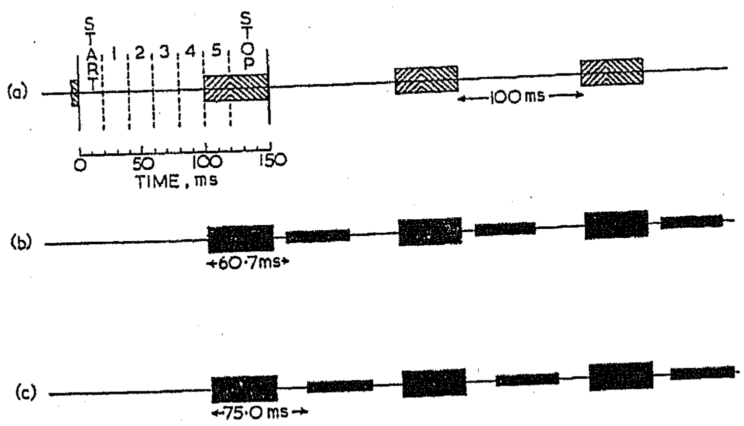


Fig. 9.—Time relationship of short- and long-route signals computed for great-circle distances.

(*a*) 50-baud start-stop 5-unit code, Letter T, test signal.
(*b*) Singapore–London, distance 10 900 km.
(*c*) Colombo–London, distance 8 750 km.

such a test signal with the 50-baud start-stop 5-unit code, which is in common use, where transmissions of a series of letter T's result in a convenient "spacing" period of 100 millisec.

Figs. 9(*b*) and 9(*c*) indicate the computed time relationship of short- and long-route signals from Singapore and Colombo respectively, neglecting factors such as heights of layers and modes of propagation; in practice, measurements suggest time delays of 61–63 millisec for Singapore and 76–78 millisec for Colombo. Measurements of the relative absorption of the signals arriving by the short and long routes present difficulties owing to the absence of precise data regarding the polar diagrams of the transmitting and receiving aerials in use and to the poor quality of signal which would result if non-directive aerials were used at each terminal. Figs. 10(*a*) and 10(*b*) indicate a typical change in the relative strength of the two signals from Singapore as received on aerials directed towards the short and long routes, respectively. Fig. 10(*c*) is a typical example of echo from Colombo.

Difficulties of reception due to backward echo may be reduced, if not eliminated, by

(*a*) Improving the front/back response of the aerials used for transmission and reception.

(*b*) Using over the short route as high a frequency as is practicable, thereby improving the ratio of short-/long-route signal strength and even eliminating the echo completely where ionospheric conditions permit the use of a frequency in excess of the maximum appropriate to the long route (see Fig. 11).

(*c*) Utilizing, during suitable periods, transmitting and receiving aerials directed over the long route.

(5) ASYMMETRY OF PERFORMANCE WITH SEASON

(5.1) Operational Results

Asymmetry of performance with season in the two directions of a given circuit has been in evidence for some considerable

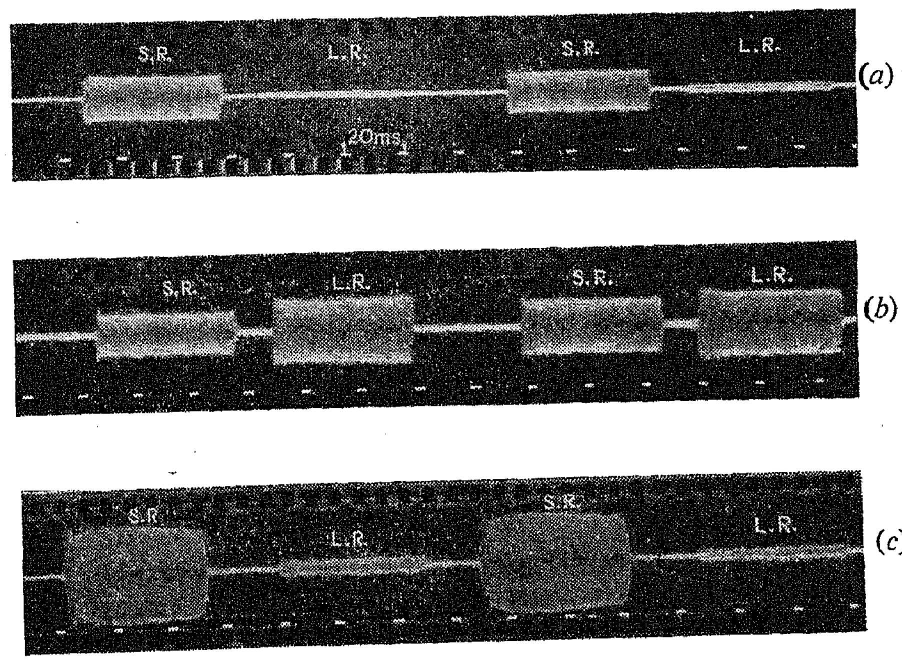


Fig. 10.—Relative receiver inputs for short- and long-route signals, reception in London and transmission from Singapore and Colombo directed in each case over the short route.

(*a*) Singapore, 17 955 kc/s, 1100 hours G.M.T., 15th November, 1952; receiver aerial directed 077° (short route).
(*b*) Singapore, 17 955 kc/s, 1100 hours G.M.T., 15th November, 1952; receiver aerial directed 257° (long route).
(*c*) Colombo, 18 545 kc/s, 1050 hours G.M.T., 30th June, 1949; receiver aerial directed 093° (short route).

S.R. Signal via short route.
L.R. Signal via long route.

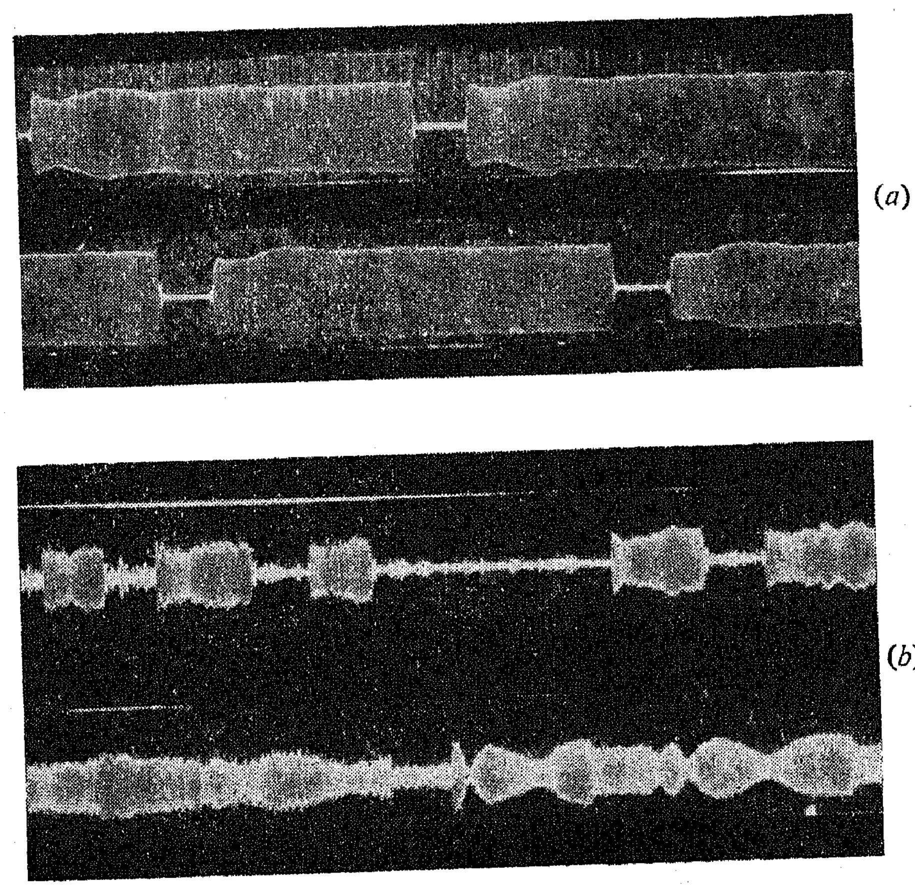


Fig. 11.—Comparison of reception on two different frequencies.

(*a*) 24 340 kc/s, negligible signal distortion.
(*b*) 20 170 kc/s, distortion due to echo, multi-path, and noise effects.
(Singapore–London, short route, 1200 hours G.M.T., 3rd May, 1948.)

time, and an account follows of the performance of two main-line radio routes—United Kingdom–South Africa and United Kingdom–Australia—each of which is operated by two independent organizations.

Even though it has not always been practicable to make comparisons over identical periods of the solar cycle, nevertheless it is considered that the effects observed are real.

(5.1.1) **United Kingdom–South Africa (London–Cape Town and London–Simonstown).**

On the London–Cape Town circuit it was apparent in 1933 that reception in London was below average in summer and above average in winter, and that the converse held in regard to reception in Cape Town. It was suggested at the time that this asymmetry might result from the movement in Africa of the zone of maximum thunderstorm activity (and therefore the source of atmospherics) to the north in summer and to the south in winter. Figs. 12(*a*) and 12(*c*) show this asymmetry for the four years ending 1944.

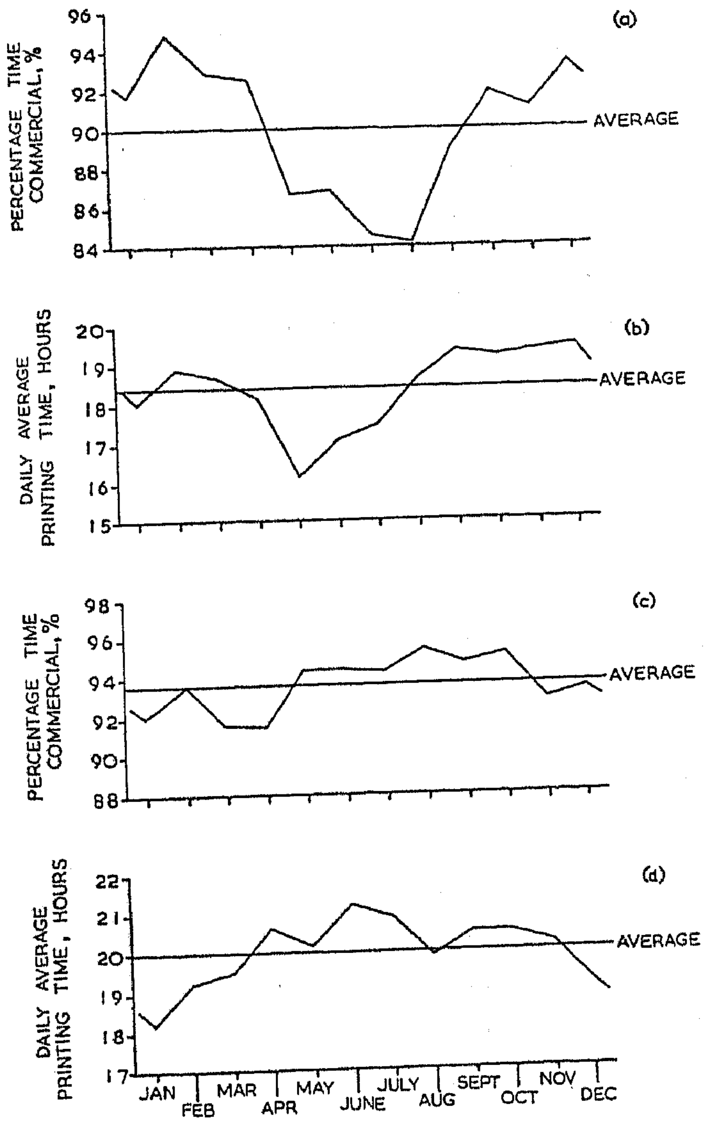


Fig. 12.—Seasonal asymmetry of performance observed by two different organizations over the same route.

(*a*) (*c*) Cape Town–London and London–Cape Town, respectively, 1941–44 (inclusive).
(*b*) (*d*) Simonstown–London and London–Simonstown, respectively, 1950–53 (inclusive).

On the London–Simonstown circuit the later introduction of teleprinter operation has disclosed similar asymmetry [Figs. 12(*b*) and 12(*d*)].

A detailed analysis of the London–Simonstown circuit on a time basis showed that on this circuit the maximum discrepancy in performance in the two directions occurred during the period 1800 to 0600 hours G.M.T. in summer, and a comparison with the London–Cape Town circuit over a number of years (see Figs. 13 and 14) showed close agreement in this respect, the effect, however, being somewhat less marked in the London–Cape Town case. The above period is known by measurement (Fig. 15) to be one for which the level of atmospherics in southern England in summer greatly exceeds that for winter.

Atmospherics, multi-path effects and scattering,[6] and interference from unwanted stations are usually more prevalent on the lower frequencies used at night (see Fig. 16), and the sharper-

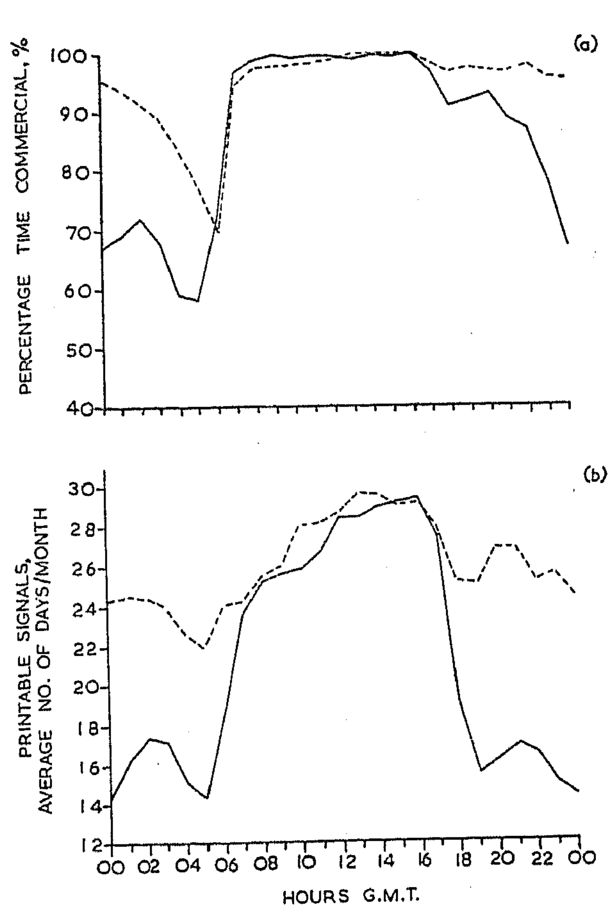


Fig. 13.—Diurnal asymmetry of performance observed by two different organizations over the same route.

(*a*) Cape Town–London circuit, summer 1940–44 (inclusive).
(*b*) Simonstown–London circuit, summer 1950–53 (inclusive).
– – – – – From London.
———— To London.

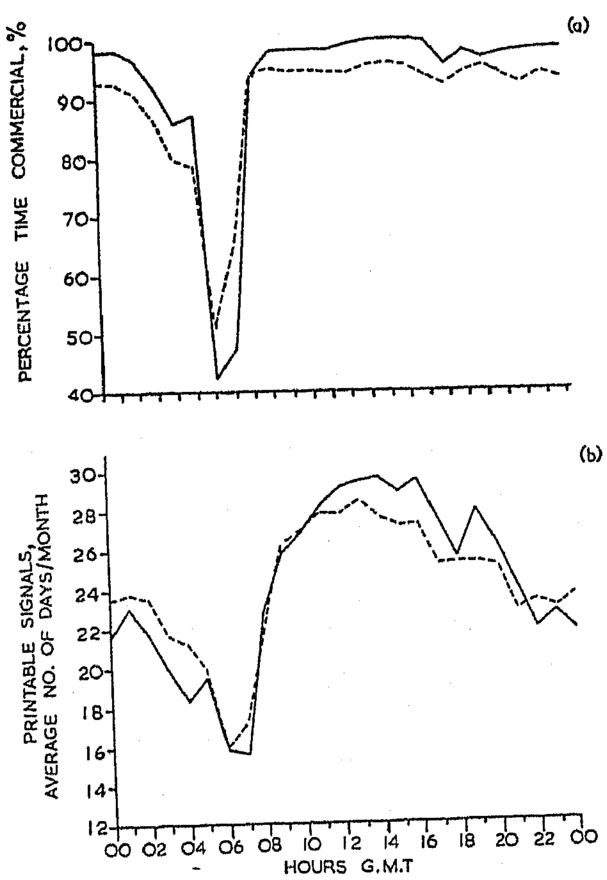


Fig. 14.—Diurnal asymmetry of performance observed by two different organizations over the same route.

(*a*) Cape Town–London circuit, winter 1939–40 to 1944–45 (inclusive).
(*b*) Simonstown–London circuit, winter 1949–50 to 1953–54 (inclusive).
– – – – – From London.
———— To London.

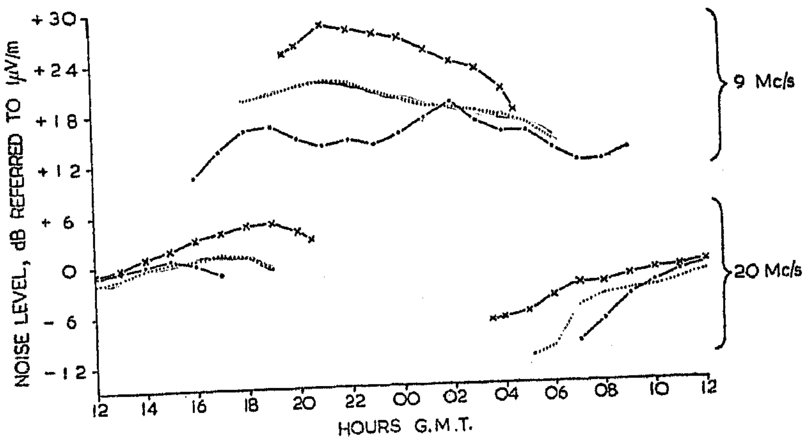


Fig. 15.—Seasonal variations in atmospheric peak noise levels. Somerton, 1939 (receiver bandwidth, 2·5 kc/s).

×—×—×—× Summer.
.... Autumn.
—·—·—·— Winter.

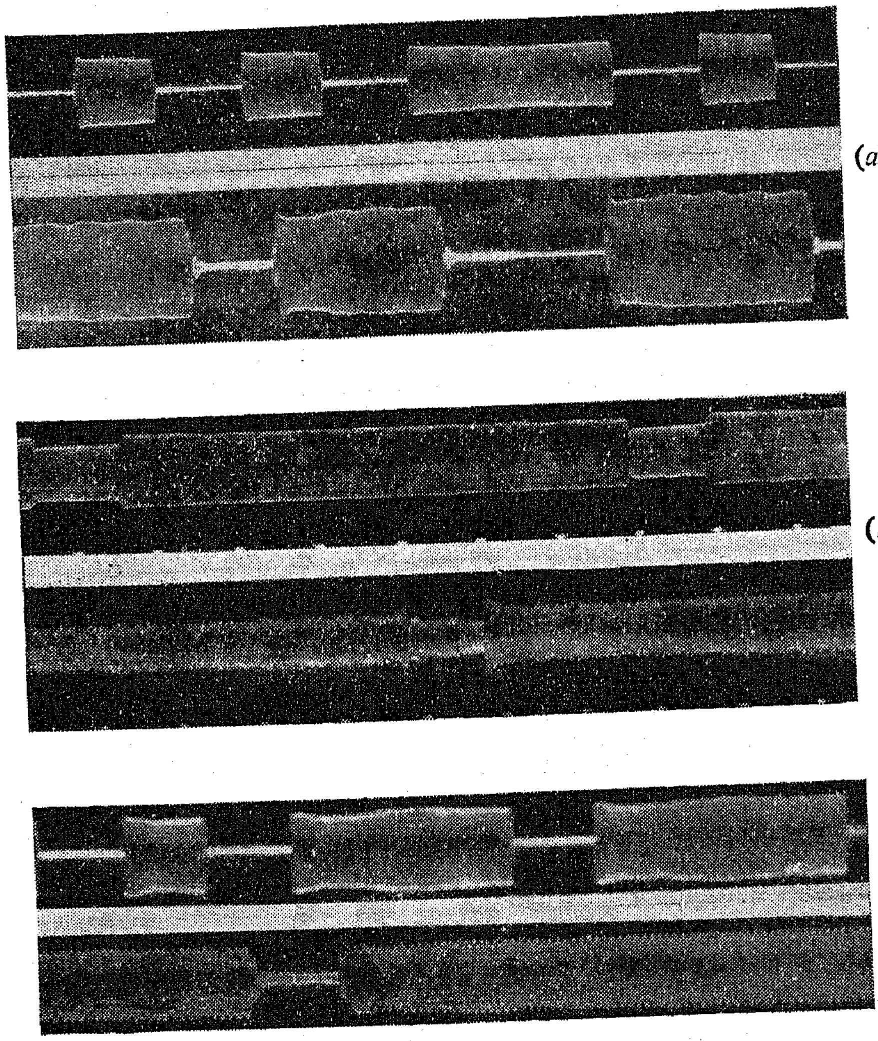


Fig. 16.—Comparison of reception conditions on high and low frequencies (upper traces and lower traces respectively).

(*a*) Simonstown–London (ICW), April, 1948.
(*b*) Halifax–London (FSK), March, 1948.
(*c*) Gibraltar–London (ICW), July, 1948.

beam aerials of the London–Cape Town circuit and the use thereon of synchronous (as distinct from start-stop) keying may account for the fact that this circuit is affected to a lesser degree than the London–Simonstown circuit.

(5.1.2) United Kingdom–Australia (London–Melbourne and London–Harman).

The United Kingdom–Australia route presents extreme difficulties owing to contrasting day and night conditions; in general no regular communication is possible, either over the short or long route, in winter from about 2300 to 0600 hours G.M.T. and to a lesser extent in summer around midday. Although

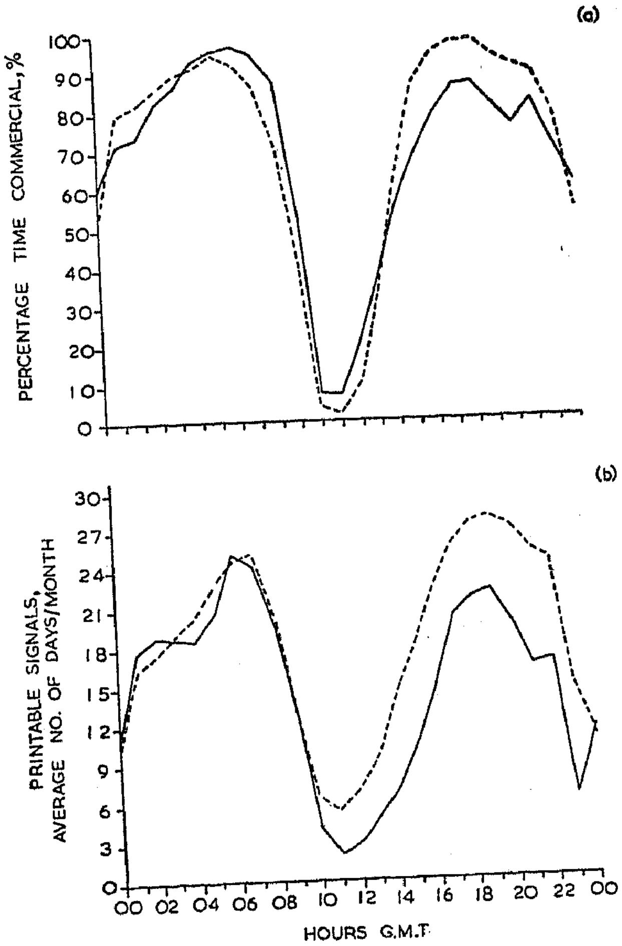


Fig. 17.—Diurnal asymmetry of performance observed by two different organizations over similar routes.

(*a*) Melbourne–London circuit, summer 1938–41.
(*b*) Harman–London circuit, summer 1950–53.
- - - - From London.
——— To London.

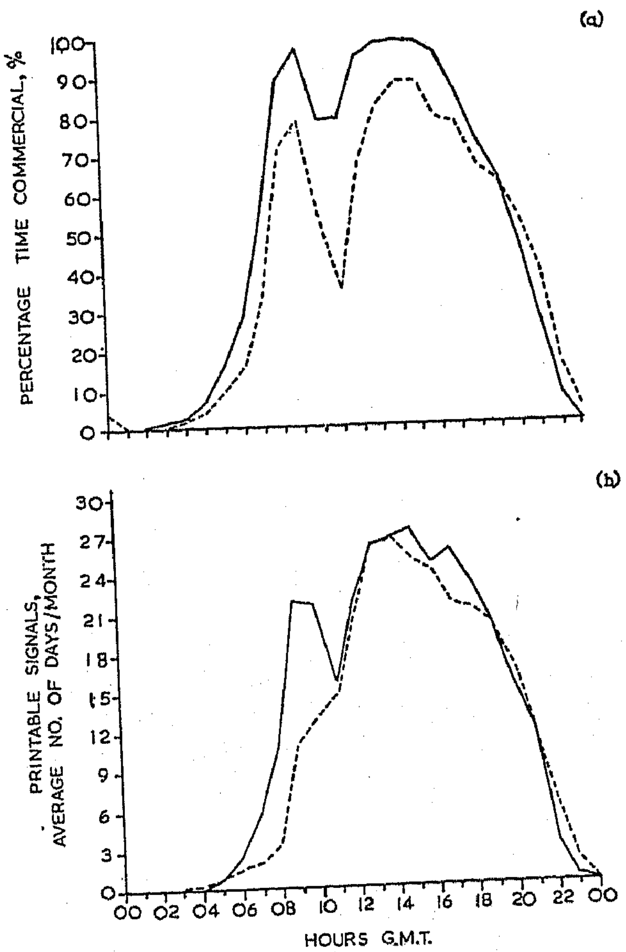


Fig. 18.—Diurnal asymmetry of performance observed by two different organizations over similar routes.

(*a*) Melbourne–London circuit, winter 1937–38 to 1941–42 (inclusive).
(*b*) Harman–London circuit, winter 1949–50 to 1953–54 (inclusive).
- - - - From London.
——— To London.

there are significant geographical differences in route from Melbourne and Harman to London, asymmetry of performance with season is very similar in many respects (see Figs. 17 and 18), reception in summer in London being inferior to that at the Australian terminals notably between midday and midnight when the short route is in use, whereas in winter, reception in London is superior for most periods of the day irrespective of the route in use. The effect was very pronounced in the sunspot-minimum period of 1944 (see Fig. 19). The difficulties at the

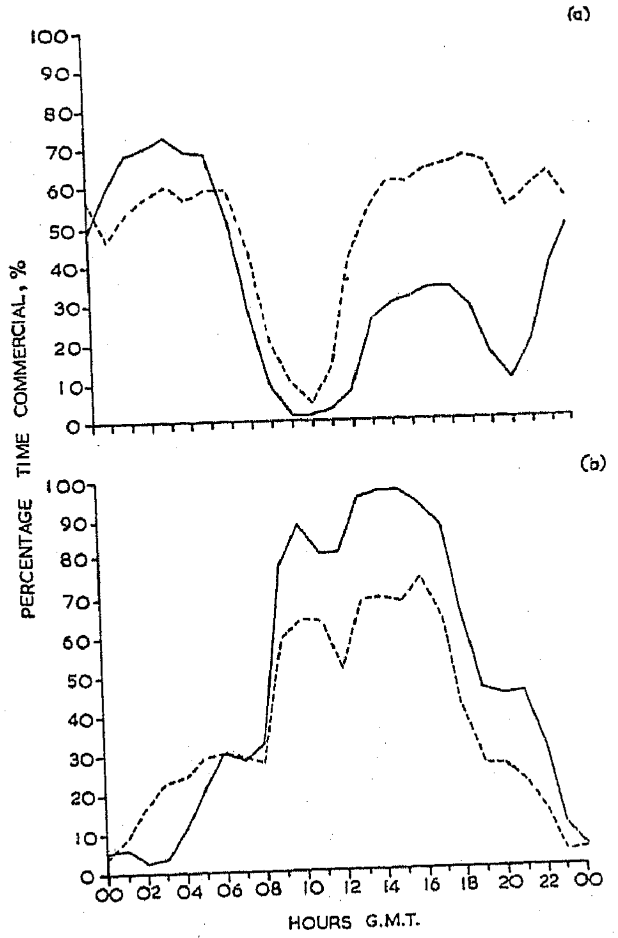


Fig. 19.—Diurnal asymmetry of performance of the London–Melbourne circuit in sunspot minimum.

(a) Summer, 1944.
(b) Winter, 1944–45.
- - - - From London.
——— To London.

London terminals in summer are usually attributed to low signal strength and to atmospherics and interference from other stations.

(5.1.3) **Other Circuits.**

Asymmetry of this type has also been observed on the London–Colombo, London–Singapore and Montreal–Melbourne circuits, and an analysis of the last-mentioned circuit covering 19 years is shown in Fig. 20. These examples suggest that such asymmetry may be characteristic of other medium- and long-distance circuits.

Unless cognizance thereof is taken, assessments of the performance of new equipment, methods of keying, etc., introduced in one direction of a given circuit, may be erroneous if based on a comparison with the performance in the other direction, in which the equipment has remained unchanged.

(5.2) Possible Reasons for Asymmetry Effects

A thorough investigation of the reasons for asymmetry effects such as those described in the previous Section has not been carried out because it would necessitate making measurements of various kinds which are not practicable in a commercial or military communication organization. There is, however, a tendency to attribute the observed effects to non-reciprocity in

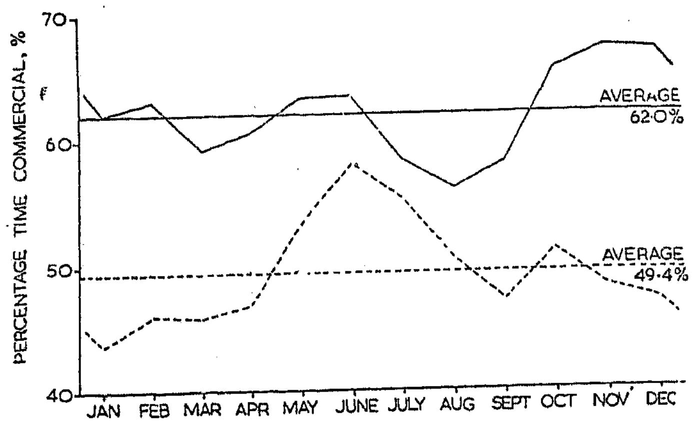


Fig. 20.—Seasonal change in performance illustrating the asymmetry of the Melbourne–Montreal circuit.

1935–53 inclusive.
——— To Montreal.
- - - - From Montreal.

the propagation of radio waves in opposite directions and a resulting difference in the signal strength at opposite ends of the circuit, even assuming the terminal equipments to be identical.

To prove the existence of non-reciprocity in the strict sense of the word would require accurate absolute measurements of the field strength to be made simultaneously at both ends of the circuit. It is doubtful whether such a measurement could be made to an accuracy of better than 3 dB, and hence it would be difficult to detect with any certainty a difference of less than 6 dB between the field strengths at the two terminals of the circuit. Consequently it would be difficult, in any particular case, to prove experimentally that long-term asymmetry was due to anything other than differences in the terminal equipment or its utilization.

It is important to recall that in some instances there is a seasonal change in asymmetry which is superposed on the long-term asymmetry. While long-term asymmetry might be explained in terms of permanent differences in the efficiency of the equipment and personnel, it is very unlikely that such an explanation would be acceptable for seasonal changes in asymmetry. It is therefore necessary to give some consideration to other possible reasons for the observed effects.

(5.2.1.) **Ionospheric Propagation.**

Long-distance circuits necessarily involve several reflections from the ionosphere. It has therefore been suggested that the asymmetrical effects just described might owe their origin to non-reciprocity in propagation through the ionosphere. It is difficult to conceive how this might occur, except possibly as a result of the directional nature of the horizontal component of the earth's magnetic field.

The theory of propagation through the ionosphere in the presence of a magnetic field has not been fully investigated because of serious mathematical difficulties. Furthermore, there is very little sufficiently well-controlled experimental evidence to throw any light on the magnitude of any non-reciprocal effects which may exist. It has recently been found by R. W. Meadows that the fading characteristics of pulse signals propagated via the ionosphere over a 700 km north-south path in the United Kingdom may, although they are usually identical, differ appreciably at the opposite ends of the circuit. Consequently the possibility of non-reciprocal ionospheric propagation cannot be eliminated, but no estimate can be given of the magnitude of any such effect which may exist.

(5.2.2) **Atmospheric Noise.**

Even if it is assumed that the received signal strengths at the terminals of a long-distance circuit are the same, there may still be a marked difference in the atmospheric noise signal appearing

at the respective receiver inputs. The consequent difference in signal/noise ratio may be large enough to lead to a difference in the performance of the circuit in opposite directions. It is not generally appreciated, however, that when directional receiving aerials are used it is not possible to predict the noise level at a receiving site by making use of the usual contour charts showing noise grade. Most of the important commercial and military circuits use directional aerials for transmission and reception, but although a gain in signal/noise ratio will always result from the use of an appropriate transmitter beam aerial, this does not necessarily follow when a beamed receiver aerial is used.

The noise level at a receiving site is usually measured using a vertical receiving aerial having substantially omnidirectional characteristics. Alternatively, in the absence of measured values, it is usual to estimate the total field due to all thunderstorm centres likely to produce an important contribution. Both procedures are equivalent to integrating the noise power received from all directions, and the total power obtained in this way is that used in producing the well-known noise-grade charts. When a directive receiving aerial is used, only those storm centres which lie in the direction of the main beam or side lobes can make an important contribution to the total received noise power. The total noise power received using a directive aerial may be greater or less than that for a non-directional aerial, but provided that the main beam of the directive aerial is correctly aligned for the wanted signal, there will be no reduction in signal/noise ratio. It is conceivable, however, that when the wanted signal lies in a direction approximating to that of a storm centre covering a fairly wide range of bearings, a small difference between the direction of arrival of the signal and the axis of the main beam might lead to a serious drop in the strength of the wanted signal with no accompanying drop in received noise power.

Unfortunately, no systematic measurements have been made of the directions of arrival of atmospheric noise as a function of time for different locations and on different working frequencies. Furthermore, it would be difficult to make reliable estimates of these from the available data on thunderstorm centres. Data from Somerton suggest that atmospheric noise may be the cause of the asymmetry observed on the United Kingdom–South Africa circuit, but in the absence of suitable data from South Africa, the evidence cannot be taken as conclusive.

(5.2.3) **Ray Paths.**

If it is assumed that the aerials at the terminals of a radio circuit are identical and that the ionospheric reflecting layers are horizontal, even if there is a seasonal change in the angle of incidence of the radiation, the effective gain of the receiving aerials will change in the same way at both terminals and no asymmetry effects will be produced. If the vertical polar diagrams are not the same, a seasonal change in the angle of incidence could lead to a seasonal change in the difference between the effective aerial gains at the terminals. Owing to the existence of tilts in the reflecting layers, it is possible that the angles of arrival at opposite ends of a circuit may be different; under these circumstances asymmetrical performance, both long-term and seasonal, could occur even if the vertical polar diagrams of the aerials at the terminals were identical.

Very little information on the angle of elevation of radiation from distant transmitters is available because of the difficulty of making reliable measurements; furthermore, no measurements are known to have been made simultaneously at both ends of a circuit. However, in view of the tilts which must exist in the reflecting surfaces along a multi-hop circuit, it is difficult to believe that the angles can be the same at both ends of such a circuit to within a few degrees. The bottom lobe of the rhombic aerials frequently used in practice is fairly wide, and a small change in the direction of arrival of a ray near the peak of the lobe would not alter the effective gain by very much. However, there is some evidence for believing that on some circuits the radiation arrives at a lower angle than had hitherto been thought likely. If such rays are, in fact, being received on the sloping lower side of the bottom lobe, small changes in incident angle could be responsible for quite large changes in effective gain and might lead to asymmetry. It is unlikely that there would be a compensating change in the effective gain for atmospheric noise, since this is probably received over a much wider range of angles of elevation.

(6) MISCELLANEOUS PROPAGATION EFFECTS

(6.1) Maximum Latitude of Ray Path

Investigations carried out notably in Japan[7] and the United States[8] have shown the advantages to be gained by a reduction of the maximum latitude of ray path during magnetically disturbed conditions.

As a result of extensive measurements of field intensities of European transmitting stations in Japan, the main transmitting and receiving installations in the vicinity of Tokyo were re-sited in 1936–37, reducing the maximum latitude of ray path, in the case of the London circuit, as follows:

(*a*) In the direction London–Tokyo from 71°34′ to 69°23′.
(*b*) In the direction Tokyo–London from 72°00′ to 70°31′.

Subsequent intercomparisons of circuit data throughout 1938 indicated that, during certain periods of magnetic disturbances, communication was confined to the direction London–Tokyo.

From similar considerations it follows that under disturbed conditions reception in the United Kingdom of transatlantic signals will be least affected in southern England, and likewise the reception in the United Kingdom of New York will be superior to that of Montreal. If the radio installations were actually in the centres of New York, Montreal and London, the maximum latitudes of ray path would approximate to 53°30′ and 55°00′ for the New York and Montreal routes, respectively.

(6.2) Propagation by M Reflections

During the period of relatively high solar activity of 1936–39 the Buenos Aires–London circuit was operated over the long route on 17 Mc/s for about four to six hours centred on 1400 hours G.M.T. throughout the equinox and winter months to obviate the effects of backward echo on the short route [Section 4.2(*c*)]. This procedure was facilitated by the existence of suitable aerial arrays at Buenos Aires and London directed on Tokyo. It is of interest to note from Fig. 21 that, from 1200

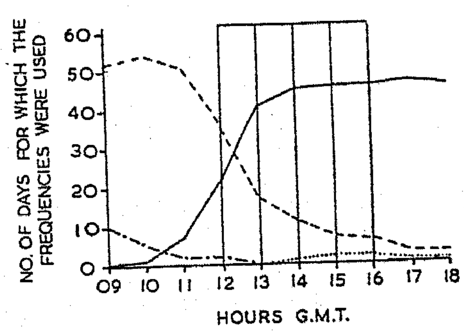


Fig. 21.—Comparison of frequency usage over the whole and part of the Buenos Aires–Tokyo–London route.

Winter, 1936–37.

Period when the frequency of 17 Mc/s was in regular use. Buenos Aires–London long route.

7 Mc/s, 9 Mc/s, 13 Mc/s, 18 Mc/s — Tokyo–London.

to 1600 hours G.M.T. in the winter of 1936–37, the Tokyo–London circuit required frequencies significantly lower than 17 Mc/s, which was regularly used on the Buenos Aires–London long route passing near to Tokyo. The similarity in geographical disposition of these two routes appears to be inconsistent with the large difference in frequency usage for the period in question. The subsequent installation of the required high-gain aerial equipment suitable for frequencies of the order of 27 Mc/s [Section 4.2(*b*)] made the use of the long route unnecessary.

Another instance of the same phenomenon can be cited for the London–Melbourne path. Between approximately 1900 and 2300 hours G.M.T. in winter, Melbourne signals on 11 Mc/s have been readable at Barbados (long route over London) at times when this frequency has been too high for the Melbourne–London section.[1]

The fact that such circuits have sections exceeding 4 000 km in length requiring frequencies significantly lower than those used over the entire circuit tends to support the use of the two-control-point method of computing maximum usable frequencies, and suggests the possibility in such cases of propagation by M reflections in the ionosphere.

(7) CONCLUSIONS

It would appear that no complete explanation of some of the effects described is possible in the absence of a better understanding of the mode of propagation between transmitter and receiver, and of the directional properties of atmospherics.

Moreover, whilst considerable improvements have been introduced in the field of terminal equipment, relatively little progress has been made in regard to the determination of optimum angles of projection and the most efficient design of receiving aerial.

By restricting transmission and reception as far as practicable to optimum trajectories, not only would the present congestion in the h.f. bands be reduced, but the quality of reception would be improved as a result of a reduction of multi-path effects and of interference from atmospherics and unwanted transmissions.

(8) ACKNOWLEDGMENTS

The authors wish to thank the Air Ministry and the Canadian Overseas Telecommunications Corporation for permission to use information regarding their circuits. They also wish to thank Miss M. B. Andrews and Mr. B. W. Smith of the Royal Naval Scientific Service for their assistance in the presentation of the data in the paper.

The paper is published by permission of the Admiralty, the Engineer-in-Chief, Cable and Wireless, Ltd., and the Director of Radio Research of the Department of Scientific and Industrial Research.

(9) REFERENCES

(1) SMALE, J. A.: "Some Developments in Commercial Point-to-Point Radiotelegraphy," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 345.

(2) CABLE AND WIRELESS LTD.: "Field Strengths required for Commercial Telegraphy on Short Waves," *Marconi Review*, 1936, No. 60, p. 10.

(3) TREMELLEN, K. W., and COX, J. W.: "The Influence of Wave-Propagation on the Planning of Short-Wave Communication," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 200.

(4) ECKERSLEY, T. L.: "Short Wave Wireless Telegraphy," *Proceedings of the Wireless Section, I.E.E.*, 1927, **2**, p. 85.

(5) WOOD, K. L.: "Empire Telegraph Communications," *Journal I.E.E.*, 1939, **84**, p. 638.

(6) ECKERSLEY, T. L.: "Multiple Signals in Short-Wave Transmission," *Proceedings of the Institute of Radio Engineers*, 1930, **18**, p. 106.

(7) AMISIMA, T., *et al.*: "Comparison of Field Intensities of Short Waves received Simultaneously from European Stations at Two Different Points in Japan," *Report of Radio Research in Japan*, 1937, **7**, p. 75.

(8) KENRICK, G. W., *et al.*: "Relation between Radio-Transmission Path and Magnetic Storm Effects," *Proceedings of the Institute of Radio Engineers*, 1938, **26**, p. 831.

DISCUSSION BEFORE THE RADIO SECTION, 9TH FEBRUARY, 1955

Mr. F. Axon: The present time is most appropriate for the presentation of these very interesting papers. We enter a new sunspot cycle in much the same spirit as we enter a new year—confident that we shall be able to make better use of the spectrum. The commencement of each cycle produces this same optimism, but as the cycle progresses we realize more and more how little we know and how difficult this problem of h.f. propagation really is, and sometimes we wonder whether we are making any real progress.

There is, however, ample reassuring evidence to show that, so far as h.f. broadcasting is concerned, the percentage of circuit reliability has improved steadily, firstly because more reliable forecasts of critical and usable frequencies are now available, and secondly because we are able to take advantage of operational experience.

The B.B.C. has a fairly good record of the effects of disturbances on the North Atlantic circuit, which correlates very closely with the findings on the general trends given in the paper by Messrs. Jowett and Evans. The main difference is one of degree. Possibly the explanation lies in the range of frequencies available to the two types of service and the relative spacing of the bands. The broadcasting service is often tied to a fixed frequency, published some weeks beforehand. For example, the broadcast service might well have been operating at around 15 Mc/s at the time of the disturbance, whereas the point-to-point service operated at around 17 or 19 Mc/s. Again in sunspot minimum the lowest band which the European broadcaster can use to serve North America is 6 Mc/s; when that is affected by disturbances there is no reserve, and therefore he could not assess the effects of using a lower frequency.

This points to the necessity for a more scientific yardstick to measure the effects of these disturbances, and perhaps there is a need for one experimental circuit operating close to the predicted m.u.f. throughout the 24 hours.

I do not think that we should go through another sunspot cycle without attempting to plan short-wave broadcasting. One of the biggest obstacles in the past to such a plan has been the lack of data of the kind given in these papers. At each successive conference we find the frequency requirements of the various countries increasing. One reason, and a very important one, is or has been the lack of confidence in predictions; another reason has been lack of confidence in the ability to forecast disturbances. The result is often the use of two or three frequencies in different bands in order to act as an insurance against faulty predictions and disturbances. For these reasons many countries require three frequencies in three separate bands. It is only when we have sufficient data of the type given in the papers that we shall be able to convince all countries that there is not much benefit from these spectrum-wasteful practices. We have evidence in the papers that even on a commercial circuit, where it is possible to change the frequency and use 2-way point-to-point communication to decide the best frequency, one can lose 30% of the

time during the peak listening period for broadcasters. If it is possible to lose 30% of the time with control of the circuit, it is obvious that it is possible to lose more when there is no control, as with broadcasting circuits, so that there is not much point in continuing to radiate on three frequencies. I hope that we shall eventually evolve a system which will close down the transmitter during a severe disturbance rather than waste valuable power.

It has been made clear by Messrs. Jowett and Evans that their paper presents a subjective analysis carried out on operational circuits. However, do the authors consider that the use of a second supporting frequency would, in fact, increase the total percentage time of reliability on this circuit, because if not, that may be quite a step forward? If, in fact, the broadcaster does not obtain any additional guarantee of percentage of time of reliability, he will realize that there is no point in using a supporting frequency.

In examining the data, which must have been fairly extensive, have the authors derived any ideas on whether the commercial circuits examined were working close to the optimum predicted frequency? I believe that it is the general practice, in broadcasting at least, to have a supporting frequency at a band below. When the circuit began to fail, was it possible to drop the frequency by 1 or 2 Mc/s and maintain the circuit, or did that have no effect?

Mr. G. Millington: Most of the propagation data used in assessing the performance of h.f. communication systems are based on vertical-incidence measurements made at a large number of ionospheric observatories. The papers are a reminder that some of our knowledge of the ionosphere, as it affects long-distance h.f. circuits, has been obtained from traffic records. Reference has been made to the fact that it is sometimes possible to receive a transmission from Melbourne at Barbados when it cannot be heard at London, and it was the consideration of such anomalies that led Tremellen to propose the control-point method of deciding the m.u.f. for such circuits. Although it is empirical, it has become internationally recognized, since no satisfactory alternative explanation exists that is based on the conversion from vertical- to oblique-incidence behaviour.

With regard to conditions at sunspot minimum, there are two effects that combine to make propagation difficult at times of magnetic disturbance, namely the low value of the m.u.f., which may approach the l.u.f. and leave no margin between for communication, and an increase in the magnetic activity itself due to M-regions on the sun in the sunspot-minimum epoch. I presume that the latter factor is primarily responsible for making communication difficult as compared with the sunspot-maximum years rather than the greater susceptibility to any particular magnetic storm.

I am not convinced that the evidence presented for non-reciprocity refers to the reciprocity theorem as usually understood in propagation theory. The readings on the graphs are given in percentage of time, and it is not obvious what they mean in terms of field strength. Moreover, does the apparent non-reciprocity refer to the m.u.f. or the l.u.f.? As the earth's field renders the ionosphere doubly refracting, it is necessary to give careful consideration to the characteristic polarizations of the ordinary and the extraordinary modes in stating the reciprocity theorem. I feel that we need to know more about the actual terminal conditions for the circuits before we can conclude that the propagation is really non-reciprocal.

With regard to the use of more than one frequency for a given circuit, what is the authors' opinion, on purely technical grounds, of the desirability of making frequent frequency changes in order to remain close to the m.u.f., as opposed to the practice of making as few changes as possible and using more power if necessary to overcome absorption difficulties?

Mr. A. F. Wilkins: During the past 30 years or so a considerable quantity of scientific information concerning long-distance h.f. propagation must have been accumulated by communication organizations. Such information, much of which has not been published, would be of great interest to those who are investigating the mechanism of propagation of high-frequency waves, and for this reason I welcome the papers.

Messrs. Humby, Minnis and Hitchcock refer to the phenomenon of reception in Barbados of signals from Melbourne which had passed unheard over London. Similar phenomena have been observed recently at the Radio Research Station while using the back-scatter technique for investigating h.f. propagation. In this technique pulse-modulated h.f. waves are transmitted from a directional aerial; after reflection by the ionosphere the waves arrive at the ground at a distant point and are there scattered by irregularities. Some of the scattered energy returns to the vicinity of the transmitter and is receivable on the same aerial. Fig. A shows a portion of an equivalent path/time record

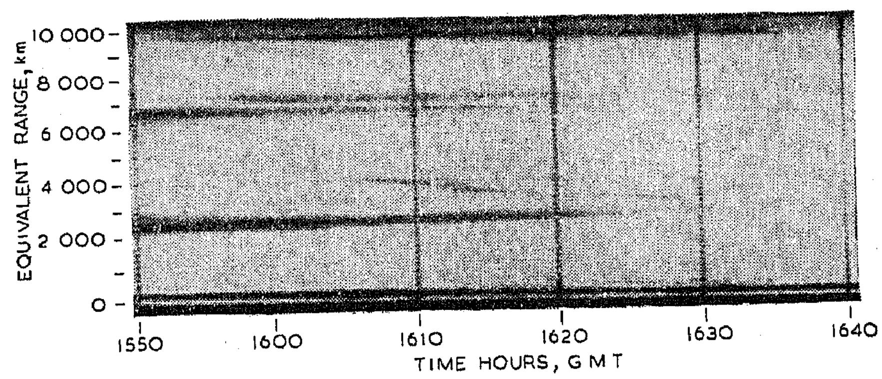


Fig. A.—Continuous range/time record of back-scattered echoes received at Slough on 17th December, 1954.

Frequency—19 Mc/s.
Aerial—Rhombic directed 142° east of north.

of back-scattered echoes in which those at about 2 000 km range are scattered from the edge of the skip zone for F2-layer transmission; the echo band at about 6 500 km range probably corresponds to scattering at the ground after a 2-hop trajectory by way of the F2-layer. It will be seen that the echo at 10 000 km range persists for some minutes after the fade-out of the other echoes, and this fact leads to the conclusion that the signals would have been audible at 10 000 km (i.e. south of Madagascar) when inaudible at shorter ranges.

As one of the important requirements for successful h.f. communication is that the F2-layer ionization density shall be adequate to reflect waves of the frequency in use, I suggest that it would have been more satisfactory if Messrs. Jowett and Evans had related the lost time to an ionospheric rather than a magnetic index. Such an index might have been derived from measurements of F2-layer critical frequencies made near the great-circle path of the signals. In this way it would have been possible to show why the lost time may be large during quiet magnetic conditions at sunspot minimum as well as during storms.

In Section 3.2.4, Messrs. Jowett and Evans suggest that a discrepancy in the variations of lost time with magnetic conditions on the Montreal–London telegraph and telephone circuits may be attributable to differences in the characteristics of the aerials used in the two systems. This suggestion might have been more convincing if the authors had added some information on how these aerial characteristics differ.

It would be of interest to know how much of the lost time on the direct circuits considered by the authors was recovered by the introduction of relay working.

Mr. H. Stanesby: The papers are very interesting and no doubt contain data of value to students of the ionosphere in the sense that all scientific knowledge is valuable. But we may ask

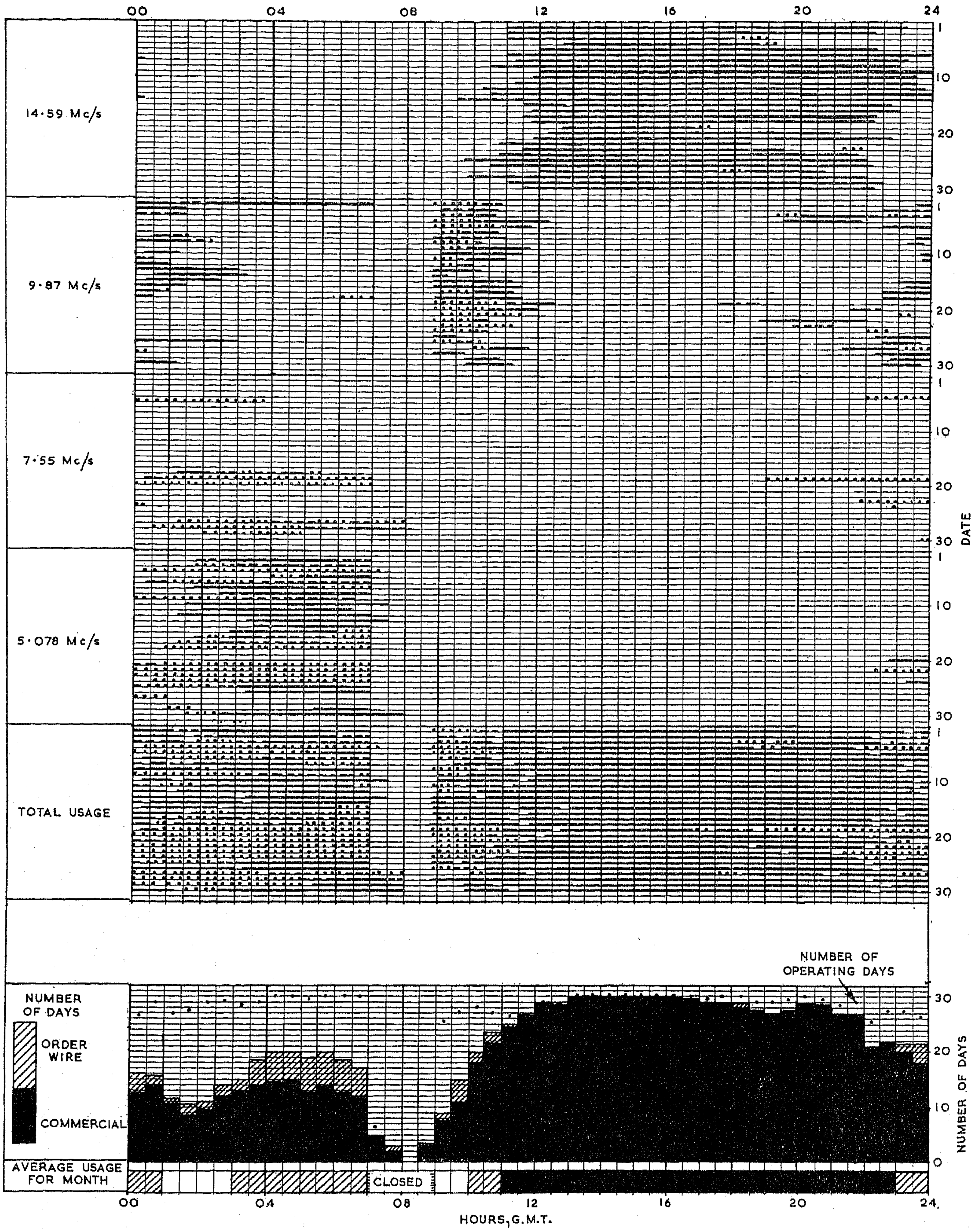


Fig. B.—Analysis of the performance of the New York–London telephony circuit (high-power system using m.u.s.a. reception) for September, 1953.

- ——— Commercial.
- — — — Order wire.
- · · · · · Unusable.
- ■ Commercial for more than 80% of the operating days.
- ▨ Commercial or order wire for more than 50%, but commercial for less than 80%, of the operating days.
- □ Commercial or order wire for less than 50% of the operating days.

whether they have any engineering value. What is the significance of a paper like that presented by Messrs. Jowett and Evans, which relates the performance of North Atlantic circuits to solar and magnetic data? First, the paper shows that we must expect the performance to vary greatly not only throughout the year, with poorest results near the equinoxes, but throughout the sunspot cycle, with worst results near sunspot minimum. This is of great value to operating organizations, because it tells them when they should mobilize their resources to deal with the most difficult conditions. Moreover, by relating circuit performance to more fundamental data like magnetic and solar activity, knowing the latter, the performance can be checked at any time, and we can see whether changes in plant and operating methods are having a beneficial effect. Papers like this are essential if long-distance radiocommunication is to be improved, and it would be helpful if wider use were made of an agreed method of representing circuit performance. A method which is coming into use on many of the circuits operated from this country is shown in Fig. B. The day-by-day behaviour on each frequency is shown for a month on a single diagram, and the performance for telephony is graded as "commercial," "order wire" or "unusable"; while for telegraphy it is graded as "slips once," "slips twice" or "unreadable" for speeds not less than 100 words per minute. The overall performance for the month is summarized in the two lower sections of the diagram.

The paper by Messrs. Humby, Minnis and Hitchcock is different, since it draws attention to a number of phenomena revealed by studies of circuit performance. Attention is drawn to the asymmetrical performance of certain radio circuits, particularly that between London and Melbourne. Our own more detailed studies confirm their findings, but show that asymmetry is virtually confined to the short route. The long route is almost completely symmetrical. We have found, with the co-operation of the D.S.I.R., that, when reception in the United Kingdom is worse than in Australia, the noise in the United Kingdom is up to 13 dB higher than that in Australia, and as the receiving aerials are orientated towards regions of high noise, we think that noise is at least partly responsible for the asymmetry. In addition, we have made carefully controlled tests to find out whether ionospheric absorption over a radio path depends on the direction of transmission, whilst ensuring that when the direction of transmission is reversed there is no change of aerial performance at either end. We did this by using the same aerial for transmission and reception. With the co-operation of the Overseas Telecommunications Commission (Australia) we carried out tests with Australia in which the direction of transmission was reversed at 2 min intervals, and we computed the mean transmission loss for each interval separately. We found that on some days transmission was substantially reciprocal; while on other days it was persistently non-reciprocal to the extent of as much as 8 or 10 dB. Substantially similar results were obtained across the North Atlantic with the co-operation of the American Telephone and Telegraph Company. The tests were of comparatively limited duration, and it happened that over both routes, when non-reciprocal conditions prevailed, the loss was usually less for signals incoming to the United Kingdom. However, it is difficult to believe that, on the average, the loss in one direction would exceed that in the other. It is to be hoped that the non-reciprocity of radio circuits will be studied further by ionospheric physicists.

Mr. H. F. Finch: It has been shown in the papers that there is a strong correlation between the losses on the transatlantic transmissions and the Abinger magnetic index, K. In Figs. 1 and 2 of the paper by Messrs. Jowett and Evans the seasonal correlation is quite marked, and there is also a correlation with the solar cycle. Another correlation, which is not specifically mentioned but which may be of interest, is in the diurnal variation. It is stated that the losses are heaviest between 2100 and 0600 hours. If we plot the mean value of K for each of the separate eight 3-hour intervals taken over a period we find a very definite diurnal variation in the form of an approximate sine curve with a maximum at about 2000 hours. The local K-indices for other stations give similar curves, which shows that there is a local time effect. The maximum varies slightly in time; Cheltenham (Maryland, U.S.A.) shows it at about 2100 hours local time, or 0200 G.M.T. This suggests that over the Atlantic the period of maximum magnetic activity lies somewhere between 2000 and 0200 hours G.M.T., which may provide a further instance of correlation of losses with magnetic activity.

Messrs. Jowett and Evans refer in their paper to a study of magnetic activity and sunspot number. The index used in this study was not K, but U, based upon the day-to-day differences in the mean value of the horizontal intensity, H. The reason that the K-index does not show the correlation to so marked a degree is that we can get the same mean K-index from a series of moderately large values as from a mixture of large and small ones. During the period around solar maximum great storms are interspersed with quiet periods. Recently, preceding the solar minimum, we have had a large number of prolonged recurrent storms in which the level of activity has not been really high, but apparently high enough to be associated with the losses. Thus, using the K-index, we get a magnetic-activity curve which, in the mean, is rather flat taken over the solar cycle but which correlates extremely well with the losses.

That losses continue after periods of severe activity, even when the K-index has resumed normal values, is of interest. Magnetic activity is usually accompanied by a depression in the mean value of H, and after a severe storm it is some days before the mean value of H is restored to normal, although little disturbance is apparent. This has been used as evidence of the existence of a ring current outside the earth at a distance of several earth's radii; but here we have evidence from radio observations that something is happening in the ionosphere which seems to coincide with the post-perturbation period when the value of H is low.

Dr. R. L. Smith-Rose: In view of the importance of the subject-matter contained in the papers it is desirable to consider exactly the manner in which the material and, in particular, the results are presented. I tried to compare a curve from one paper for a particular circuit with a curve for a similar circuit in the other paper, and I found that they were 180° out of phase. One paper refers to the characteristics of circuit performance, while the other emphasizes the facts relating to lost time.

I merely wish to emphasize the importance of agreeing on some uniform and easily-assimilable manner of presenting the data. The matter is not merely of national importance, it is bound to arise in international discussions, and it would be very creditable if we set an example in the presentation of material.

Messrs. Humby, Minnis and Hitchcock mention the importance of studying the angle of arrival of radio waves, and if possible, taking advantage of this knowledge, while in the paper by Messrs. Jowett and Evans, one at least of the circuits refers to reception on a steerable array. What improvement is obtained by the use of the m.u.s.a. system?

Can Messrs. Humby, Minnis and Hitchcock explain the difference in attenuation of the round-the-world echoes mentioned in Section 4.1? Apparently the first set of signals passes round the world with an attenuation of 13 dB, and on the second time round it suffers an attenuation of only 3 dB. Is this a general result, and is there some simple explanation, or alternatively is it a curious phenomenon which may deserve more detailed investigation later?

Mr. H. Leigh: In both papers, suggestions are made as to how

the percentage commercial time might be increased, and it may be of interest to know how the Post Office have followed up two of these points on one of the North Atlantic telephony systems.

The great-circle path from New York to London passes uncomfortably close to the northern auroral zone, and we have, since 1952, made use of an experimental relay to divert the route southwards, using plant which was available in the Cable and Wireless, Ltd., station in Barbados. The time lost on the direct route is mainly during winter nights; at times of low sunspot activity the night-time performance drops from about 90% commercial time in summer to below 20% in winter, despite the fact that the system is equipped with high-power senders at both ends and has m.u.s.a. reception. The relay has low-power senders at three of the four stations, and the aerials have not been specially designed, but we have succeeded in getting roughly twice the commercial time from it during winter nights compared with the direct circuit, and even greater commercial time would have been obtained had we been able to find interference-free frequencies.

Messrs. Jowett and Evans make the suggestion that an aerial for the North Atlantic route should be designed for a maximum efficiency at a vertical angle of 15° or less, and this, at the winter night frequency of 5 Mc/s, is somewhat difficult to attain with a rhombic aerial, but in 1951 we tackled this problem by building a large transmitting rhombic aerial with a 600 ft side and 140 ft high. This aerial has a maximum gain at 5 Mc/s of some 8 dB over the normal rhombic aerial used, and at low angles, a relative gain considerably larger than this. In 1954, the American Telephone and Telegraph Co. followed suit and built two such

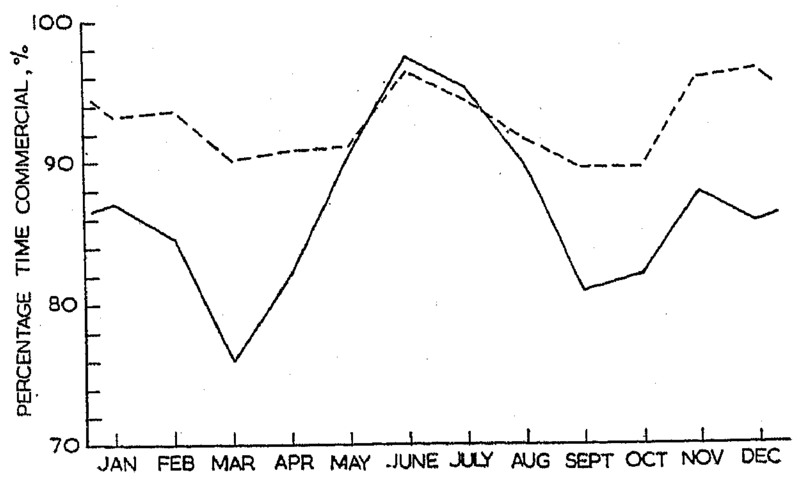


Fig. C.—Seasonal change in performance of the Montreal–London telegraph circuit for the period 1933–52.

– – – – – 1937, –38, –39, –46, –47, –48, –49 (Sunspot number—85 or above).
———— 1933, –34, –42, –43, –44, –45, –52 (Sunspot number—35 or below).

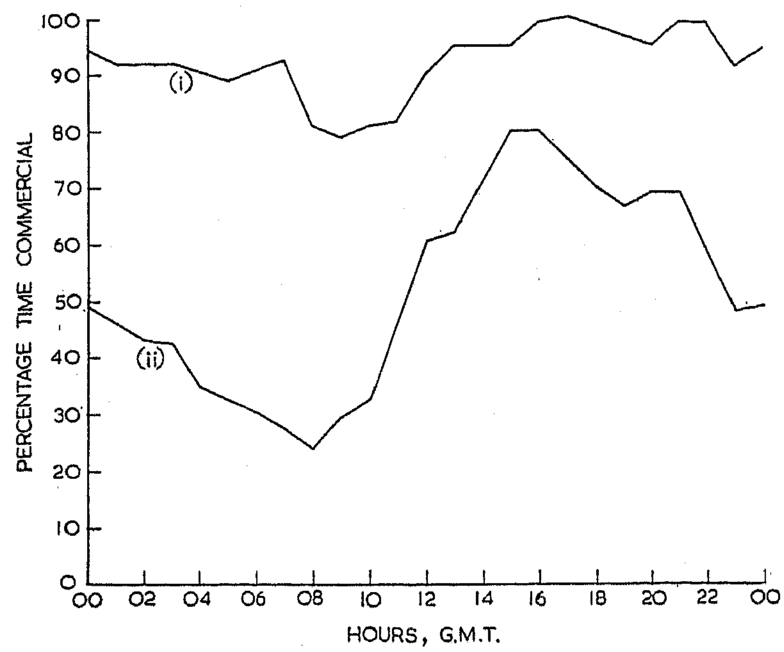


Fig. D.—Typical diurnal characteristics of the Montreal–London telegraph circuit.

(i) August, 1940. Abinger *C*-value = 0·77.
(ii) March, 1941. Abinger *C*-value = 1·15.

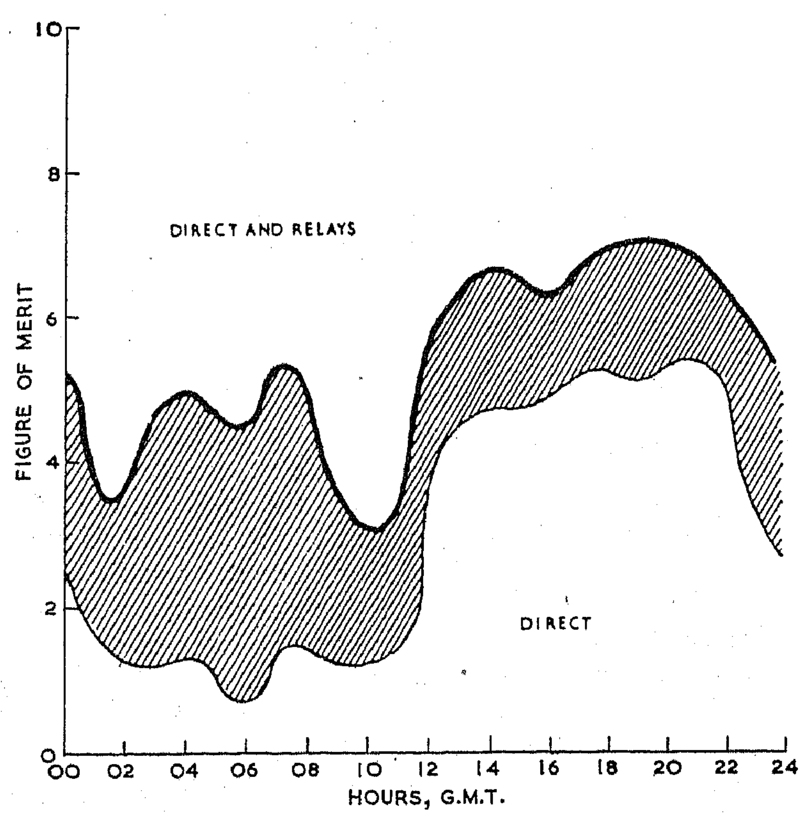


Fig. E.—Diurnal performance of the Montreal–London radiotelegraph circuit and associated relays (Ascension and Barbados) under disturbed ionosphere conditions.

(i) Mean performance during disturbed days February, March, April and September, 1946.
Mean sunspot number = 77.
Mean Abinger *C*-value = 1·65.

(ii) Mean performance during disturbed days September, October and November, 1953.
Mean sunspot number = 12·
Mean Abinger *C*-value = 1·17

rhombic aerials, one for transmission and one for reception; these aerials have a 1 000 ft side and are 100 ft high. A series of measurements at 5 Mc/s made in March, 1954, showed that, with a low-angle transmitting rhombic aerial in the United Kingdom working to a low-angle receiving rhombic aerial in the United States, the receiver input voltage at midnight G.M.T. averaged some 24 dB, and at 0400 hours G.M.T. reached over 30 dB—higher than that of a similar system with normal-sized rhombic aerials. Averaged over a large number of observations the gain was about 27 dB. This is a very considerable gain and has resulted in a marked improvement in performance during the present winter. Some of this improvement is no doubt due to improved propagation conditions, but I think that at least half of it may be attributed to the new aerials.

Mr. J. A. Smale: Messrs. Jowett and Evans state that it is during sunspot minimum, and particularly at night time in winter, that by far the greatest lost time occurs. Fig. C shows the seasonal change in performance of the Montreal–London telegraph circuit for a considerable number of years of high and low sunspot activity. It will be seen that at both levels of solar activity the most difficult times are at the equinoxes and not in the winter. Perhaps Messrs. Jowett and Evans mean that the period centred on the winter is worse than that centred on the summer.

In Section 4.1 the period 2100–0600 hours G.M.T. is mentioned as being generally the most difficult from the point of view of transatlantic radiocommunication—an opinion which may result from the closing of the Montreal–London telephone circuit at 0600 hours G.M.T. Fig. D illustrates typical diurnal characteristics of the telegraph circuit and shows that the worst period of the day is centred on approximately 0730 hours G.M.T. and is not the period 2100–0600 hours G.M.T.

I think that this has some effect on Fig. 2 and Section 3.2.4, in which it is pointed out that, whereas in 1950–51 and 1951–52 the telegraph circuit had a pronounced dip in lost time in the winter, the telephone circuit does not appear to have this effect. I think it is likely that, if the whole 24 hours had been taken into account and the early morning added to the results, the dip would have been very prominent in the telephone results as well, and there would be no need for an effort to find an explanation based on the difference in aerials.

Mr. Leigh has given some very interesting figures of time recovered on the New York–London radiotelephony circuit by the use of a relay at Barbados. Fig. E shows the very marked effect which can be obtained by the use of relays on the Montreal–London telegraph circuit. It is of particular interest in that it illustrates the difference in the performance and value of relays under different conditions of sunspot and magnetic activity.

Mr. A. M. Humby: The opinion expressed by Mr. Smale (and illustrated in Figs. C and D) that the greatest difficulties in regard to the operation of the Montreal-to-London telegraph circuit occur in spring of sunspot-minimum years and after 0600 hours G.M.T., is equally applicable, in my view, to other transatlantic radiotelegraph circuits, e.g. from New York and Halifax to London.

In this connection Fig. F illustrates the performance of the

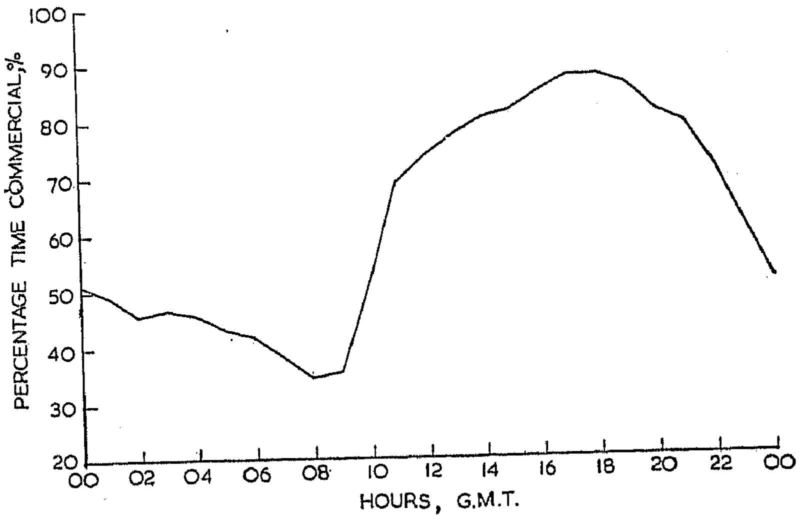


Fig. F.—Performance of the Admiralty Halifax-to-London teleprinter circuit for the spring months (February, March and April) of the years 1950–54 (inclusive).

Admiralty Halifax-to-London teleprinter circuit for the spring months of the years 1950–54 (inclusive), i.e. years of relatively low sunspot activity.

As will be seen, the spring period from about 0600 to 0930 hours G.M.T. was one of extreme difficulty—a fact which would seem to conflict with statements in the Summary and Section 4.1 of the paper by Messrs. Jowett and Evans.

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Messrs. J. K. S. Jowett and **G. O. Evans** (*in reply*): Messrs. Axon and Millington both raise the question of whether the circuits considered were working close to the optimum predicted frequencies. Within the limits set by the restricted frequency complements, the circuit operator does tend to follow fairly closely the m.u.f. curves for the North Atlantic circuits. This is more particularly true, of course, during the winter night-time conditions at and near sunspot minimum when the greatest difficulties in circuit operation usually arise. Mr. Axon also asks whether the use of a supporting frequency would improve circuit reliability. We do not, in general, consider that the reliability would be improved, but under disturbed conditions it might be useful to bring the day-to-night transition frequency into use an hour or so before the predicted fade-out time of the day frequency. This would give the broadcaster some protection against the early fade-out of the normal day frequency. The data for 19th September in Fig. B illustrate this point.

In reply to Mr. Millington, the factor primarily responsible for making communication difficult around sunspot minimum is the disturbed ionospheric conditions shown by the high degree of magnetic activity. In other words, in the absence of ionospheric disturbances the circuit time lost need not be high even under sunspot minimum conditions when there is only a small margin between the m.u.f. and l.u.f. This has been confirmed from a study of recent records on a transatlantic circuit.

Mr. Wilkins suggests that we should use an ionospheric index rather than a magnetic index. The difficulty here is in selecting the most appropriate ionospheric observatory. We have frequently had severe interruptions on North Atlantic circuits when there has been little evidence in the Slough records of an ionospheric storm. Mr. Wilkins's queries on the use of relays are covered by the comments of Messrs. Smale and Leigh.

In Fig. B Mr. Stanesby has shown a shadowgraph form of analysis which is now being adopted, not only in this country, but also by some Commonwealth countries, and we share his view that a generally agreed method of representing circuit performance is desirable. The development of such analyses goes some way towards meeting Dr. Smith-Rose's plea for a uniform manner of presentation of such data. However, there is a distinct advantage, in some forms of analysis, in referring to lost time rather than to commercial time. In further reply to Dr. Smith-Rose, we think that the use of the m.u.s.a. system of reception effects a reduction of lost time on the New York–London circuit of about one-third.

The evidence that Mr. Leigh has put forward indicating the need for transmission at low angles at low frequencies to cover winter night conditions at sunspot minimum is of considerable interest. We are grateful also to Mr. Finch for drawing attention to the differences between the various magnetic indices; his remarks on the local time variations of magnetic activity suggest further lines of investigation of the relation of circuit interruptions to magnetic disturbances.

In reply to Mr. Smale, we agree that the Montreal telegraph circuit shows the greatest amount of lost time at the equinoxes, but, for the two telephone circuits considered, the most difficult period is centred upon winter and includes the equinoxes. Messrs. Smale and Humby question the statement that 2100–0600 hours G.M.T. is the most difficult period for transatlantic communication, although their own Figs. D and F show a very marked decline in performance from approximately 2100 hours G.M.T. The period centred on 0730 hours G.M.T. is certainly very difficult but is a normal feature of the circuit under quite ionospheric conditions, as indicated in curve (i) of Fig. D. Furthermore, this period shows regular seasonal variations which are probably due to the very low m.u.f. at the western end of the circuit just before dawn rather than to the disturbed conditions which have been the main subject of the paper.

Messrs. A. M. Humby, C. M. Minnis and **R. J. Hitchcock** (*in reply*): In our explanation of the terminology used, we have been careful to distinguish between non-reciprocal propagation and asymmetrical circuit performance. As Mr. Millington rightly suggests, any violation of the reciprocity theorem could only be demonstrated in terms of measured field strengths. The data which we have presented in Section 5 all refer to asymmetry in the operational performance of certain circuits as measured by, for example, percentage time commercial. The seasonal changes in asymmetry, to which we have drawn attention in Figs. 12 and 20, show that it is unlikely to be due simply to less efficient operation of the equipment in one direction. Consequently we make no claim to have proved that the reciprocity theorem is invalid in the cases quoted; our evidence shows, however, that one or more causes, which have not yet been identified, but which may include non-reciprocity, must exist to account for the observed differences in performance.

On the other hand, the special tests mentioned by Mr. Stanesby appear to depend on the measurement of absolute field strength at both ends of a circuit. Assuming that the accuracy is sufficiently high and that various precautions have been taken, such tests ought to throw some light on the possible magnitude of any departure from reciprocal propagation. It is, of course, not yet possible to relate a given difference in field strength to a corresponding change in the commercial performance of a circuit. Consequently the practical significance of the difference quoted by Mr. Stanesby cannot easily be assessed. It seems likely, however, that an 8–10 dB change in field strength could be an important factor in determining the performance of a circuit operating well below the m.u.f.

Mr. Finch has suggested the interesting possibility of correlating the diurnal changes in magnetic activity with changes in circuit performance. We are not aware that any tests of this kind have been attempted, but the idea probably warrants investigation. A possible difficulty may be that, since the properties of the quiet ionosphere, and therefore of circuit performance, already contain very prominent diurnal components, it might be difficult to identify any residual effect due to magnetic activity. However, the whole question of expressing the percentage time commercial of a circuit in terms of the operating frequency and its relation to the m.u.f. and l.u.f., the level of magnetic activity, the route and possibly other factors is one which will require study if progress is to be made beyond the present simple predictions of m.u.f. and l.u.f.

Dr. Smith-Rose has drawn attention to the smaller attenuation to which forward echoes are subject on their second circuit of the earth. While some speculative explanations of this effect have been put forward, it is probably safer to regard it as a "curious phenomenon" until further study becomes worth while. It is interesting to note that the 10 000 km echo reported by Mr. Wilkins (see Fig. A) is stronger than either of the shorter-distance ones, and it seems certain that both forward echoes and back scatter, which arrive apparently without an intervening ground reflection, must be propagated by some mechanism which is not yet understood.

With regard to Mr. Millington's final question, even though it may be theoretically desirable to keep close to the instantaneous m.u.f.—assuming that its value were known at the time—practical considerations must always limit the number of frequencies used on any one circuit.

621.396.93

Paper No. 1810 R
Feb. 1955

SOURCES OF ERROR IN U-ADCOCK HIGH-FREQUENCY DIRECTION-FINDING

By K. C. BOWEN, B.A.

(*The paper was first received* 24*th November, and in revised form* 21*st December*, 1954. *It was published in February*, 1955, *and was read before the* RADIO SECTION 2*nd March*, 1955.)

SUMMARY

It is well known that errors in high-frequency direction-finding using U-Adcock systems arise from a number of independent causes. Experiments have been carried out to assess their relative importance, and statistical methods have been used to calculate the respective contributions to the total variance of error in a series of typical check bearings on known transmitters. The variance component, normally attributed to wave-interference errors, has been studied particularly, and the contributions from observational error and the effect of ground resistivity variations over the d.f. site have been estimated.

LIST OF SYMBOLS

d = Aerial spacing between the aerials of a pair in a U-Adcock direction-finder, m.
l = Aerial length for a U-Adcock direction-finder, m.
r = Amplitude of the reflection coefficient at the ground.
ϵ_r = Dielectric constant.
θ = Bearing error, deg.
λ = Wavelength, m.
ϕ_1 = Phase lag of the reflection coefficient at the ground.
ϕ_2 = Phase lag of the signal induced by the reflected wave on that induced by the direct wave for a perfectly conducting earth.
ϕ_3 = Phase lag of the signal induced by the combined direct and reflected waves on that induced by the direct wave.
ϕ_{3N}, ϕ_{3S} = Values of ϕ_3 along east–west lines through the north and south aerials for an idealized site.
ψ = Angle of elevation of the arriving wave.

(1) INTRODUCTION

The purpose of the paper is to indicate briefly the statistical nature of the errors which occur in high-frequency direction-finding using U-Adcock systems, and thereby to suggest a means of analysing the total variance of bearing error without isolating the several components and considering them separately. The study made was associated with specific bearing-classification and position-fixing problems arising in the naval high-frequency direction-finding organization, but certain of the findings are of more general interest.

The use of check bearings on known transmitters in establishing the performance of a high-frequency direction-finder is a widely accepted technique. It has been found, however, that two problems present themselves:

(*a*) There is difficulty in arranging a programme in which the radio frequencies, ranges, azimuths, propagation paths, transmitter powers and other parameters are adequately sampled and distributed similarly to those which will define the normal task of the direction-finder.
(*b*) Unless great care is taken to hide the identity of the transmission from the operator, both as regards its true bearing and its identity with previous transmissions, considerable operator bias will occur.

The methods by which these problems are overcome will generally be peculiar to the task with which the d.f. organization is concerned. It may be that (*a*) will require the deliberate setting-up of a transmitting system, although careful selection of known transmitters may obviate this; (*b*) can be overcome either by careful organization of the way in which the check-bearing tasks are fed to the operator, or by a system of random scale displacements at the direction-finder.

In the results quoted in the paper it can be assumed that the effects of operator bias are negligible and that the performance figures refer to average observational sampling times of 30 sec on medium-powered c.w. transmissions from fixed stations, in the frequency band 3–20 Mc/s, and at ranges of 700–4 000 km. Variances, given in degrees squared, are considered to be representative of an average U-Adcock d.f. site in the United Kingdom operated by an average (i.e. not highly-skilled) operator.

(2) STATISTICAL NATURE OF H.F. D.F. ERRORS

(2.1) Sources of Error

Errors which occur in high-frequency direction-finding arise from the following causes:

(*a*) Observational (operator error).
(*b*) Instrumental (errors resulting from imperfect aerial balance and similar causes; they will include errors due to the site close to the direction-finder and due to polarization errors).
(*c*) Distant site errors (due to re-radiations from objects in the vicinity of the site outside the area controlled by calibration).
(*d*) Lateral deviation in the ionosphere.
(*e*) Wave interference.

These errors, in a series of observations on the same transmitting station, can be reduced by various averaging processes, apart from that inherent in the time averaging which takes place during a single observation.

(2.2) Space Sampling

D.F. stations situated in groups, i.e. two or more stations separated by about a mile or so from one another, can be considered as being in the same geographical position for all practical purposes, but the errors occurring at them in observations on the same transmission will be substantially different. In general, the only major component of error, which will not be reduced by an averaging of the separate errors, will be that due to lateral deviation. Over the short intervals of time and space concerned in such observations, this is the same. The rapid fluctuations in this deviation, which take place owing to the irregularities in the reflecting medium, can be considered to be effectively removed by the time-averaging process inherent in a single observation, which is here defined by the observing period of about 30 sec.

A simple analysis of variance of sets of observations by stations of a group can therefore provide a measure of the correlated error variance, which is that due to lateral deviation. There are, however, two provisos that must be made.

First, there must be no possibility of operator collusion on any communication channel between the stations of a group. Any observation made by an operator is the result of a complex assessment of a rapidly-varying bearing indication, and knowledge of another station operator's result will undoubtedly introduce a bias in the form of an undesired correlation.

Mr. Bowen is at the Admiralty Signal and Radar Establishment.

Secondly, the small variance of error due to the residual octantal error, assuming that the major error from this source has been removed by calibration, must be allowed for, as this is not, in general, reducible by any form of sampling. It will be eliminated if the stations of the group are divided into two equal sections, the aerial systems of which differ in orientation by 45°. Otherwise the error component must be allowed for by a separate calculation (see Section 2.3).

(2.3) Time Sampling

It is assumed that the check-bearing programme includes repeated observations on transmissions from the same source, repeated at intervals of many hours or days, long enough for lateral-deviation errors to be totally uncorrelated. Assuming that the radio frequency of such transmissions is unchanged and that the mode of propagation is similar, errors due to instrumental and distant-site errors will be correlated and all other errors will be totally uncorrelated (the instrumental errors here include residual octantal errors).

Again a simple analysis of variance will enable the correlated variance to be assessed.

Changes in the angle of elevation and the azimuth of the arriving wave due to different ionospheric reflection heights and lateral deviations will, however, affect the accuracy of this calculation, as the required instrumental and distant-site errors in the sets of observations will not be wholly correlated in practice.[1] The variance of error attributed to instrumental and distant-site errors will therefore be an underestimate, and some empirical allowance must be made for this.

(2.4) Radio-Frequency Sampling

Observations on transmissions from the same source on different radio frequencies and at different times may be considered to be wholly uncorrelated except for the residual octantal errors. These errors will tend to increase as the angle of elevation increases and as the radio frequency increases. However, the angle of elevation tends to decrease as the radio frequency increases, and it may be considered to a first order of accuracy that the errors will remain appreciably constant.

The radio frequencies can be considered to be different in the sense that other instrumental and distant-site errors will be uncorrelated if the radio-frequency difference[1] is as much as 10%.

The analysis of variance in this case can therefore assess the order of variance due to residual octantal errors.

(2.5) Mean Errors

In the above analysis only the variance of error has been considered. It may be argued that, as most d.f. error distributions are not normal but leptokurtic, considerably more analysis is required to define them statistically. In most practical cases, however, the leptokurticity is small. In addition, the distributions are artificially curtailed so that the assessed error variance does not include "wild" errors, since these would be eliminated by a practical plotting system. It is therefore considered adequate to define a d.f. error distribution by its variance of error and its mean error.

Mean errors due to the instrument can be disregarded, as they will be discovered by the calibration procedure, and the faults can be corrected or the errors allowed for. The possibilities of mean errors arising from site or operator characteristics will be considered separately later. The only other cause of mean error is believed to be that of inadequate sampling.

The study of mean errors has not yet yielded a satisfactory method of dealing with them, as their cause is usually difficult to determine. In the type of analysis detailed above they must be taken into account when a number of different stations and operator results are put together, in order to arrive at average-variance estimates. In plotting, it is recommended that unless analyses indicate that the mean error estimated from check bearings can be applied directly as a correction in practice, the mean-square error should be used as the best estimate of variance and its inverse used as a weighting factor in plotting.

(3) ESTIMATES OF VARIANCE

(3.1) Lateral Deviation Errors

The average variance assessed for lateral deviation is about 2·0, which is reasonably in accord with figures calculated by Ross in experiments using pulse transmitters.[2]

(3.2) Distant-Site and Instrumental Errors

The variance assessed for distant-site and instrumental errors (an average over all sites at which observations were made) was about 1·2, which (as expected) is rather less than the normally accepted figure of about 2·0. The residual octantal error variance, which was very small and was estimated as 0·25, is included in the instrumental-error variance. Statistically it is not possible to separate the two sources of error; Smith and Hopkins[3] assess both the distant-site and instrumental error variance as 1·0 under conditions fairly similar to those considered here.

(3.3) Other Errors

The remaining variance is that which is normally attributed to wave-interference errors. In the observations considered, however, this figure, instead of being of the order of 1·5 (a representative figure based on less complex experiments), was about 5, after a variance of 0·8 had been added back into the distant-site and instrumental error variance.

Consideration of this large discrepancy led to an investigation of observational errors which may be expected to be statistically similar to wave-interference errors, and it has been found that, in general, a difference of 5 in variance exists between the best and worst operators at any station. In the series of observations on which this figure is based, all other sources of error could be considered to have been randomized by taking the analyses over periods of one year. Repeated yearly samples confirm these performances and also show that individual operators introduce their own peculiar mean errors into the system—sometimes as much as 1° or more.

Assuming that the best operators introduce zero variance of error,[3] the average variance component due to observational error in the naval d.f. organization is assessed as 2·5. This leaves 2·5 to be accounted for by wave interference and other causes.

(4) THE EFFECT OF IRREGULAR GROUND RESISTIVITY

(4.1) A Special Case

The study of the contribution of variations in ground resistivity to h.f. d.f. errors arose from an attempt to explain a large mean error (+1·4°) in terms of the measured resistivity characteristics of one particular site. These characteristics were rather striking (see Fig. 1) and could be considered, in a rather simplified form, as contours of constant resistivity running in straight lines east–west and with resistivity increasing from north to south from about 5 to 27 kilohm-cm. These resistivities are the average measured resistivities down to a depth of 5ft.

It seemed worth while making some calculations for such an idealized site. Any errors due to the different resistivities would be zero for waves arriving from the north or south, from considerations of symmetry, and equal but opposite in sign for

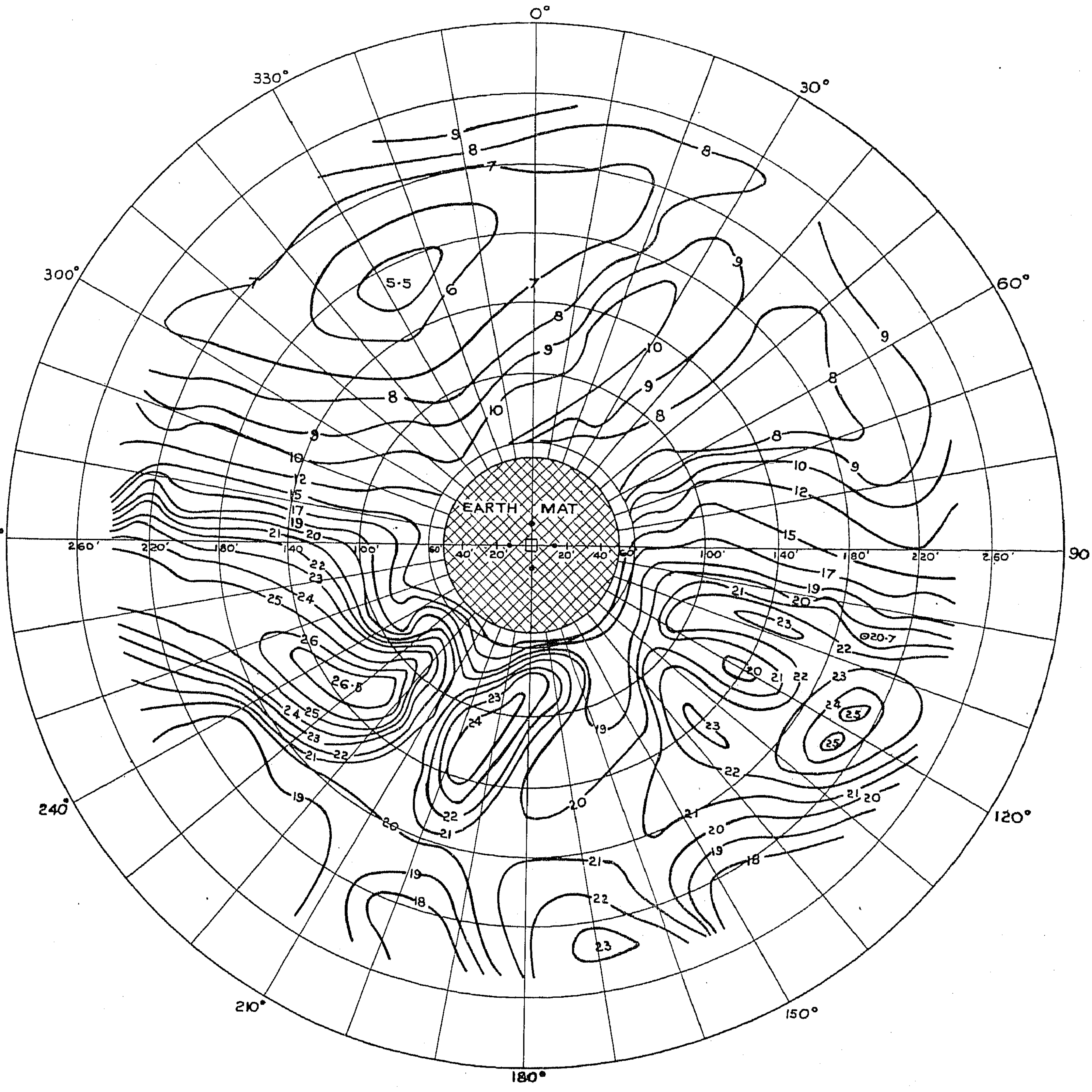


Fig. 1.—Earth resistivity in kilohm-centimetres at a depth of 5ft for one particular site.

waves arriving from the east and west. A series of observations concentrated into a limited sector might well provide a systematic error.

(4.2) Theory

Considering a single plane wavefront, if r and ϕ_1 are the amplitude and phase lag of the reflection coefficient at the ground, the phase lag, ϕ_3, of the signal induced into one aerial by the combined direct and reflected waves relative to that induced by the direct wave is given by

$$\tan \phi_3 = \frac{r \sin (\phi_1 + \phi_2)}{1 + r \cos (\phi_1 + \phi_2)}$$

ϕ_2 being the phase lag of the signal induced by the reflected wave on that induced by the direct wave, for a perfectly conducting earth. ϕ_2, obtained by integrating the signals induced into elements of the aerial along its length, is given by

$$\tan \frac{\phi_2}{2} = \frac{\sin \psi \sin \dfrac{2\pi l}{\lambda} - \sin \left(\dfrac{2\pi l}{\lambda} \sin \psi\right)}{\cos \dfrac{2\pi l}{\lambda} - \cos \left(\dfrac{2\pi l}{\lambda} \sin \psi\right)}$$

where l is the aerial length, λ is the wavelength and ψ is the angle of elevation of the arriving wave.

On the idealized site considered in Section 4.1, a plane wavefront arriving from the east will give rise to a bearing error, θ, given by

$$\tan \theta = \frac{\sin \frac{1}{2}(\phi_{3N} - \phi_{3S})}{\sin \left(\dfrac{\pi d}{\lambda} \cos \psi\right)}$$

where d is the aerial spacing, and ϕ_{3N}, ϕ_{3S} are the values of ϕ_3 along east–west lines through the north and south aerials.

Using the appropriate values of ground resistivity for this site, assumed to be 12 and 15 kilohm-cm through the north and south aerials respectively, θ has been calculated for waves arriving from the east, for various values of λ and ψ. l and d are known, and ϕ_1 and r can be calculated. This latter calculation was carried out from curves given by Terman,[4] but while these were in suitable form, the scale was very compressed in the ranges required, and some additional graphical interpolation was needed. The dielectric constant ϵ_r was taken as 10, but it is not very critical. It was assumed that the signal amplitudes at all four aerials were the same; r changes slowly with resistivity, and the relative variations made only a second-order difference in the estimate of bearing error.

Table 1

Errors due to Ground-Resistivity Variations on an Idealized Site (Section 4.1) for Waves arriving from the East

Radio frequency	Angle of elevation, ψ			
	5°	10°	15°	30°
Mc/s	deg	deg	deg	deg
5	$+3\cdot0$	$-0\cdot5$	$-0\cdot5$	$-0\cdot75$
10	$+2\cdot7$	$+1\cdot5$	$-0\cdot3$	small negative
15	$+2\cdot0$	$+2\cdot0$	$+1\cdot3$	small negative

Table 1 gives the values of θ for $\psi = 5°$, $10°$, $15°$ and $30°$, and the radio frequencies of 5, 10 and 15 Mc/s. In the cases where θ is negative, the interpolations on the curves used proved difficult, but the values are all small. The absolute accuracy of these estimates of θ is in any case unimportant, since, apart from the assumption of an idealized site, it has been assumed that resistivity is uniform with depth of penetration and no account has been taken of the fact that reflections will be scattered from a zone.

It is clear that such an idealized site could give rise to a positive systematic error if a limited sector round 090° were considered.

(4.3) Practical Considerations

In considering the performance of the d.f. station concerned, while repeated measurements of the site resistivity confirmed those on which the above arguments are based, the mean error did not remain of the same order; in fact, after a time it was found to have become negative. A study of other sites and their mean errors showed no obvious correlation. It does not seem, therefore, that the resistivity variations are likely to be a major contribution to the mean error.

Again, considering Table 1, the contribution to variance which will be included in that attributed to the instrument is not likely to be very large for single-mode reception, considering that the figures given are taken from a rather extreme case. However, when multi-mode reception is experienced the various modes will individually have different phase delays, and wave-interference errors will be caused. The errors will be similar to those caused by a combination of signals which are laterally deviated by amounts equal to the errors arising due to the resistivity variations.

In general, it would seem that the rate of change of ground resistivity over the site is of more importance in assessing the error of this type than the average resistivity, which would affect the instrumental component of variance. An earth mat changes the effective resistivity characteristics and tends to produce circular contours on an otherwise uniform site. It is thought that this will produce errors of an octantal type, although the order of these has not been assessed.

A further complication, the importance of which is difficult to assess, is the existence of vertical resistivity gradients. Slopes of as much as 5° have been indicated by measurements made.

(4.4) An Estimate of Error Variance

It would be unwise from the above arguments to give a figure for the average variance of error attributable to ground-resistivity variation. It is clear, however, that on many sites where large variations in resistivity have been measured appreciable error must arise, and that, while some of this may be seen as a mean error, it will mainly contribute to the variance of error.

The variance of $2\cdot5$ estimated in Section 3.3 for wave-interference and other statistically similar errors (observational error having been separately allowed for) must therefore be assumed to include these errors.

(5) CONCLUSIONS

The use of sampling with respect to space, time and radio frequency can enable estimates to be made of the components of variance which arise from different causes, provided that the check-bearing programme which is used is carefully designed and controlled. The estimates will not be of a high order of accuracy, but coupled with the knowledge gained from other studies, they should enable any source of large errors to be determined, and also provide useful data for certain bearing-classification and plotting systems.

The very large component of variance due to different observational abilities is disturbing, and further investigations are being made. The effect of poor sites has been superficially examined and deserves more attention.

(6) ACKNOWLEDGMENTS

In the studies on which the paper is based the author has been greatly helped by discussions with many colleagues in his own and other establishments. In particular, he is indebted to those who organized the check-bearing programmes and collated the data, and to the teams which carried out the measurements of ground resistivity on the several d.f. sites.

This paper is published by permission of the Admiralty.

(7) REFERENCES

(1) Ross, W.: "Site and Path Errors in Short-Wave Direction-Finding," *Journal I.E.E.*, 1947, **94**, Part III, p. 108.

(2) Ross, W.: "Lateral Deviation of Radio Waves Reflected at the Ionosphere," Department of Scientific and Industrial Research, Radio Research Special Report No. 19 (H.M. Stationery Office, 1949).

(3) Smith, S. B., and Hopkins, H. G.: "HF Direction Finding: Comparison of C.W. and Pulsed Transmissions," *Wireless Engineer*, 1954, **31**, p. 1.

(4) Terman, F. E.: "Radio Engineers' Handbook" (McGraw-Hill, 1943), Fig. 26(*b*), pp. 702–3.

[The discussion on the above paper will be found on page 550.]

621.396.11 : 551.510.535

Paper No. 1713 R
Aug. 1954

SOME ASPECTS OF THE RAPID DIRECTIONAL FLUCTUATIONS OF SHORT RADIO WAVES REFLECTED AT THE IONOSPHERE

By E. N. BRAMLEY, M.Sc.

(*The paper was first received* 11*th March, and in revised form* 20*th May*, 1954. *It was published in August*, 1954, *and was read before the* RADIO SECTION 2*nd March*, 1955.)

SUMMARY

Measurements have been made of the rapid directional fluctuations observed on single-component first-order ionospheric echoes, at vertical and oblique incidence (range 700 km), and on frequencies of 2·5–10·5 Mc/s. Under quiet ionospheric conditions the r.m.s. angular deviations are of the order of 1° or less, but they are appreciably greater under ionospheric-storm or spread-F conditions. The fluctuations are uncorrelated at an interval of five seconds. It is shown that a layer of irregular ionization-density within region E would account for the quiet-condition F-layer results, and this would require random variations of the phase path through the region of the order of only 3 m at a frequency of 5 Mc/s. The E-layer results appear to be affected by partial reflections at different levels within the region.

LIST OF SYMBOLS

x, y, z = Cartesian co-ordinates.
r = Distance between two points in general.
d = Distance apart, in x-direction, of two points on the ground.
a = Scale of ionospheric irregularities.
ξ = Distance apart, in x-direction, of two points in the ionosphere.
λ, ω = Wavelength and angular frequency of radio wave.
t = Total thickness of irregular region.
μ = Refractive index.
ϵ_0 = Permittivity of free space.
c = Speed of light in free space.
e, m = Electronic charge and mass.
N = Electron density.
N_0, n_0 = Mean value and standard deviation of N in the irregular region.
$\omega_0 = 2\pi \times$ critical frequency of irregular region.
ν = Electronic collisional frequency.
k = Absorption coefficient.
ϕ = Excess phase shift produced by irregularities, in traversal of thickness t by wave.
ϕ_1, ϕ_2 = Values of ϕ at two points separated by ξ in the x-direction.
ϕ_0 = R.M.S. value of ϕ.
κ = Excess absorption produced by irregularities in traversal of thickness t by wave.
κ_0 = R.M.S. value of κ.
$g = \kappa/\phi$.
$\phi_g = \phi_0(1 + g^2)^{\frac{1}{2}}$.
u = Variable of integration.
s = An interval in general.
v = Variance in general.
β_2 = Kurtosis of a frequency distribution.
erf = The error function: $\operatorname{erf} x = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^x e^{-x^2}dx$.
E = Real amplitude of wave emerging from irregular region.
E_0 = Value of E in absence of irregularities.
$\mathbf{E}$ = Complex amplitude of wave emerging from irregular region.
θ = Deviation from mean direction in an angular distribution.
$P(\theta), P_1(\theta)$ = Angular power densities in distributions.
P_0 = Total power in angular distribution.
ρ, R = Correlation functions.
b_1 = Signal/noise ratio after one passage of wave through irregular region.
b = Signal/noise ratio after two passages of wave through irregular region.
Φ_0 = R.M.S. phase difference at two points on ground.
σ = R.M.S. fluctuation in direction of arrival of wave at ground.

The M.K.S. system of units is used in this paper.

(1) INTRODUCTION

It is now well known that when a high-frequency radio wave is received after a single reflection by the ionosphere, it undergoes changes in its apparent direction of arrival. In general, these changes can be divided into two classes according to the rates at which they occur. There are also characteristic amplitude changes associated with both classes. The slower variations in direction have periods of the order of twenty minutes, and are believed to be caused by ionospheric tilts.[1] The more rapid fluctuations have periods of the order of a few seconds or less, and are associated with the short-period fading of the signal.

The rapid directional and amplitude fluctuations, as observed with a direction-finder free from polarization error, may be ascribed to wave interference effects arising from changes in the ionosphere. Thus the field incident at the receiving aerial cannot be regarded, even instantaneously, as a single plane wave, but must be looked upon as an aggregate or cone of waves travelling in different directions. This is true even of a single magneto-ionic component of the downcoming wave. In oblique-incidence transmission an appreciable contribution to the vertically polarized component of the total signal may be made by energy re-radiated horizontally from objects distributed around the receiving site. Thus variations in the incident wave appear in modified form at the receiver, resulting in one form of site error in direction-finding. A further source of the rapid fluctuations is the interference produced when ordinary and extraordinary rays are received together, with comparable amplitudes. Severe fading effects are usually produced, and directional measurements may be similarly affected. The fluctuations occurring with a single magneto-ionic component are the subject of the present paper.

In a previous paper,[2] an analysis was given of the effects produced at spaced receiving aerials by an incident angular distribution of waves. A theory of the generation of such an angular distribution by ionospheric irregularities is given in an appendix to the present paper, and by means of this and the previous work, the experimental observations are related to the parameters of the ionospheric irregularities.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

The results to be described were obtained from a series of directional measurements made with a wide-aperture spaced-loop phase-comparison system[3] known to be free from polarization error. The aerial spacing was 100 m. The experimental arrangements were as previously described,[1] the observations being made during the period 1949–1951 on single-component first-order reflections of pulse-modulated transmissions, either at nearly vertical incidence or over an oblique path of 700 km in an approximately north–south direction. The vertical-incidence measurements, which, using horizontally polarized waves, were substantially free from site errors, were used to obtain the directional deviations resulting from the angular spread or coning of the downcoming wave itself. To avoid unwanted effects produced by interference between the two magneto-ionic components, the analysis was restricted to the results obtained on occasions either when they were sufficiently separated in time of arrival to be studied independently, or when only one component was effectively present. The latter condition could arise when the other component was strongly absorbed or when it had penetrated the layer.

(2) EXPERIMENTAL RESULTS

The experimental data consisted of sequences of "snap" measurements of the direction of arrival of the signal, on various fixed frequencies in the range 2·5–10·5 Mc/s. The observations, which were photographic, were normally made at regular intervals (30 sec, except in a few special cases of more rapid recording). The rare moments of very low signal strength giving wide deviations in the apparent direction of arrival were rejected, being few enough not to affect seriously the overall results.

In the vertical-incidence results, the direction of arrival was expressed in terms of the angular deviations from the vertical in the north–south and east–west directions, while at oblique incidence (700 km) the bearing and elevation angles were used. Only the bearing fluctuations have been analysed, however, the angles of elevation being too much affected by site errors.[3]

Analysis of the magnitude of the rapid directional changes is complicated by the presence of the slower variations. In practice, a smooth curve representing the latter is drawn by eye through the measured values plotted against time, and deviations of the individual points from this curve are taken to represent the rapid fluctuations. The minimum period of oscillation allowed to occur in the smooth curve is about two minutes. Fig. 1 shows an example of this process. These results refer to measurements at nearly vertical incidence (Slough to Theale, 36 km apart in the east–west direction), and show the observed directional variations projected on to north–south and east–west vertical planes. (The average east–west component of zenith angle differs from zero on account of the east–west separation of the transmitter and receiver.) It is seen that fluctuations of the order of ½° are superimposed on the larger and slower changes, assumed to be due mainly to ionospheric tilts. Any results which were considered to be capable of widely differing interpretations as to the division into "rapid" and "slow" variations were not used.

(2.1) The Magnitude of the Fluctuations

In all cases the magnitude of the fluctuations will be expressed in terms of the r.m.s. angular deviation σ, over a period of 30 min, of the individual measurements from the smooth curve representing the slower variations in the direction of arrival.

(2.1.1) Vertical-incidence Results.

At vertical incidence the deviations in the north–south and east–west directions were nearly always found to be very similar

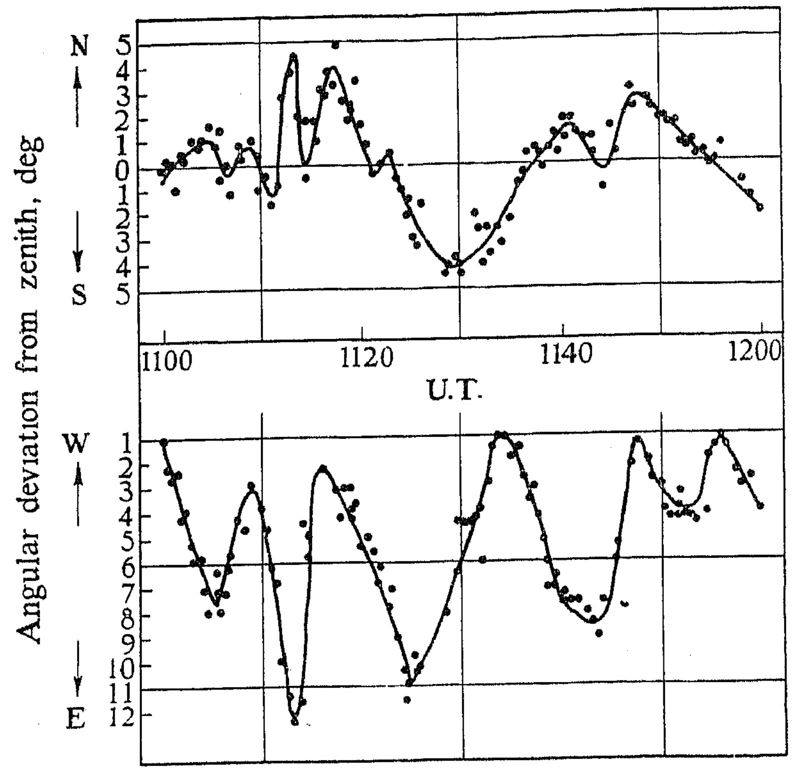


Fig. 1.—Fluctuations in direction of arrival at nearly vertical incidence.

Transmitter: Slough.
Receiver: Theale.
Date: 13th December, 1949.
First-order F reflection: 4·86 Mc/s.

in magnitude, so in the analysis they are treated together. Table 1 summarizes the results obtained for either north–south or east–west component in daytime, for ionospherically quiet days, the data being grouped in convenient frequency ranges. The mean value of σ and its estimated standard deviation are given, together with the number of periods on which the results

Table 1

R.M.S. Angular Fluctuations (North–South or East–West Component): Vertical Incidence, Daytime

F layer			E layer		
Frequency	Number of periods	σ	Frequency	Number of periods	σ
Mc/s		deg	Mc/s		deg
3	6	1·23 ± 0·08	2·5–3·0	7	2·6 ± 0·3
4·5–5·5	23	0·39 ± 0·02	3·5–4·5	10	0·70 ± 0·13
7–8	14	0·15 ± 0·02	6–8	3	0·33 ± 0·04

are based. In the E-layer results, those in the lowest frequency range were obtained on reflections from normal E, while the rest refer to the sporadic-E layer. The results for region F refer chiefly to the F2 layer, but include some reflections from F1 at the lower frequencies.

For both layers, these results show that the magnitude of the fluctuations is roughly inversely proportional to the square of the frequency. It is also noticeable that, on the lower frequencies, the spread of the values of σ derived from individual periods is greater in the E-layer than in the F-layer results.

A few results have been obtained under ionospheric storm conditions, characterized by abnormally low critical frequency, and great height, of the F2 layer. On most of these occasions the above values of σ for quiet days are markedly increased, usually by a factor of about two. The effect is most clearly seen in the F-layer results, but is also indicated in those for the E layer.

F-layer results taken during the night show a much greater

range of values of σ on different occasions than is obtained from daytime experiments. On frequencies of 4–5 Mc/s, for example, values of σ as low as $0\cdot 2°$ and as high as $2\cdot 4°$ have been recorded for the F-layer in summer time. It appears that on quiet nights the results are similar to those obtained by day. The larger values were obtained on occasions when the condition known as "spread F" was recorded in routine vertical-incidence ionospheric measurements at Slough. At Slough this condition, which occurs most commonly in the early hours of the morning before sunrise, is comparatively infrequent in midsummer (up to about five nights per month being affected) but at other times of the year it is more common, occurring on about 90% of midwinter nights. Thus while low values of σ, such as those obtained during the day, may often be obtained for the F-layer during summer nights, much higher values, of the order of 3°, may be expected in winter.

A few summer night-time observations on sporadic-E echoes on $2\cdot 5$ Mc/s have yielded values of σ between 2° and 3°, which are similar to the daytime normal-E-layer results on this frequency.

(2.1.2) **Oblique Incidence Results.**

Table 2 gives the mean values of the r.m.s. rapid bearing fluctuations observed on the 700km path in quiet daytime conditions.

Table 2

R.M.S. Bearing Fluctuations: Oblique Incidence, Daytime

F layer			Normal E layer		
Frequency	Number of periods	σ	Frequency	Number of periods	σ
Mc/s		deg	Mc/s		deg
5–6	20	$0\cdot 60 \pm 0\cdot 04$	5–6	25	$0\cdot 58 \pm 0\cdot 05$
$7\cdot 5$–$8\cdot 5$	22	$0\cdot 41 \pm 0\cdot 05$	$7\cdot 5$–$8\cdot 5$	8	$0\cdot 51 \pm 0\cdot 11$
$9\cdot 5$–$10\cdot 5$	46	$0\cdot 20 \pm 0\cdot 02$	$8\cdot 5$–10	12	$0\cdot 54 \pm 0\cdot 11$

The F-layer results thus again show the systematic fall in σ with increasing frequency, but no clear dependence on frequency is seen with the E-layer results. The comparatively wide spread of values of σ for the E layer is again noticeable.

The oblique-incidence results again show a tendency for larger values of σ to occur during ionospheric storms, but this is not at all clearly marked on the E echoes. In this case the higher values are chiefly due to wave-interference effects on occasions when marked splitting of the echo was observed. It seems likely that such splitting is due to stratification or large-scale patchiness of the ionization in the E region.

Of oblique-incidence measurements at night, only a few have been under conditions where the ordinary and extraordinary (F-layer) echoes were separable, and these, made in summer, yielded results similar to those obtained by day, as given in Table 2. On other occasions, much higher values of σ, up to 2°, were obtained, particularly in winter, when spread-F echoes were usually shown on either or both of the vertical-incidence recorders situated at each end of the trajectory. Similarly, night-time sporadic-E-layer observations have given values of σ as low as $0\cdot 5°$ but more commonly 1–2°, the higher values usually occurring when there were signs of more than one component in the echo.

(2.2) Distribution of Angular Deviations

The distribution of the individual values about the smooth curve representing a given set of observations is found to be symmetrical and approximately normal in shape, but a detailed analysis (restricted to vertical-incidence results, to avoid the effect of site errors) shows a tendency for it to be leptokurtic, or peaky, compared with the true normal distribution. The index of kurtosis,* β_2, was estimated for eighteen typical sets of results, for F- and sporadic-E-layer reflections, and values between $3\cdot 5$ and $5\cdot 5$ were obtained. For a normal distribution, $\beta_2 = 3$. Although the values obtained are not very precise, each being based on a sample of 100–150 values, a definite leptokurtic tendency is indicated by the results, which also show the characteristic curvature of such distributions when plotted on arithmetic probability paper. The theoretical distribution of apparent directions of arrival indicated by a phase-comparison system such as that actually used, when receiving a continuous angular distribution or cone of waves, has been calculated [Reference 2, eqn. (35)], and it may be shown that for a narrow cone, this distribution, while symmetrical and bell-shaped, is highly leptokurtic. It is made more nearly normal by the presence of a specularly reflected component along with the continuous distribution, a condition which is often realized, particularly in the daytime.[2,4]

(2.3) The Time Scale of the Fluctuations

The directional variations under consideration have hitherto been referred to as "rapid" or "second to second" fluctuations. The actual rate of variation has been studied by taking measurements more frequently than the usual 30-sec intervals. A number of sequences of observations have been made in daytime, at intervals of 5 or 10sec, and the rate of fluctuation examined by calculating the autocorrelation coefficient between successive readings at these intervals. In every case the value of this coefficient was found to be insignificantly different from zero, a result which applies to E- or F-layer echoes, at either vertical or oblique incidence. The same conclusion was also drawn from a sequence of vertical-incidence measurements at night, on F-layer echoes. This result means that observations made at 5-sec intervals may be regarded as independent, and thus if time averaging of the fluctuations is required the readings may be taken at least as rapidly as once every 5sec without appreciable duplication of information.

A few sets of more rapid bearing observations have been taken (at ¼-sec intervals) on pulse transmissions at oblique incidence, using a simple ciné camera and an Adcock direction-finder with cathode-ray-tube bearing indication. These have shown the existence of bearing fluctuations with periods of a few seconds, and hence appreciable correlation between bearings taken at intervals of up to one or two seconds may exist. More observations would be required to put this matter on a statistical basis. It may be noted that in bearing observations on continuous-wave signals, where the fluctuations are mainly due to interference between different modes of propagation, the results are generally uncorrelated at an interval of between one and two seconds.[5]

(2.4) The Distance Correlation of the Fluctuations

A number of sets of observations have been made on the rapid fluctuations, at vertical and oblique incidence, using two identical direction-finding systems 27km apart, normal to the direction of arrival of the signals. The individual photographic measurements were synchronized at the two stations to within less than a second by telephonic communication, and in the later stages of the work this was improved by the provision of a land line connecting the stations. An impulse, generated by the

* β_2 is defined as the ratio of the fourth moment about the mean to the square of the variance.

firing of the camera at one station, was passed along the line and automatically fired the camera at the other station, without appreciable delay. In all cases a correlation coefficient insignificantly different from zero was found, a result to be expected (as shown in Section 6.5) from the observed magnitude of the fluctuations and the separation of the receivers.

(3) DISCUSSION OF RESULTS

(3.1) The Nature and Location of the Ionospheric Irregularities

The fluctuations in direction of arrival which have been described are clearly due to changes occurring in the irregularities of ionization distribution in the upper atmosphere. It is clear that these irregularities must be situated at or below the height of reflection, i.e. about 125km in the case of E echoes. Irregularities are known to exist in the E region, in the nature of a variable stratification or cloudiness of the ionospheric structure, and random variations in ionization density in the lower part of the region have been assumed to be responsible for v.h.f. transmission over long distances.[6] Since the v.h.f. signals are obtained at all hours of day, it appears that the irregularities are always present, even when the ionization density is insufficient for the usual h.f. reflections from the region to be seen. It seems likely that random irregularities in this region may similarly be responsible for the directional fluctuations on E echoes described in the present paper. In the case of reflections from the F layer, the fluctuations might be caused by these irregularities in region E, or by irregular structures in the F layer itself. The former effect is suggested, though not proved, by results obtained when E and F echoes are studied at the same frequency, over the same or closely succeeding periods of time, since there is a tendency for a small angular spread on one echo to be associated with a small spread on the other, and similarly for larger spreads. Also, comparatively large spreads on F reflections have often been noted when E echoes are present at the same time. Again, if the irregularities were distributed throughout a large range of heights in the F region, at which waves of various frequencies are reflected, the marked fall of σ with increasing frequency would scarcely be expected since each increase in frequency would involve a longer path through the region of irregular ionization.

In the Appendix is outlined a theory of the scattering produced when a plane wave passes normally through a region containing irregularities in ionization density. The refractive index is assumed to be near unity, and the effects of variations in both refractive index and absorption, arising from these irregularities, are included. The analysis is thus an extension of that given by Jones, Millman and Nertney,[7] who consider irregularities varying sinusoidally across the wavefront. The scattering model considered remains two-dimensional, however; i.e. variations in ionization density are assumed to occur vertically, in the direction of propagation, and horizontally except in one particular direction for which the density at a given height is taken to be invariant with horizontal distance. That is to say, at any instant, the surfaces of constant density are cylindrical, so that the resulting angular distribution of energy, of which the parameters are calculated, represents a "fan" rather than a "cone" of waves.

In interpreting their v.h.f. results,[6] Bailey *et al.* used an analysis of scattering given by Booker and Gordon,[8] which includes the effect of three-dimensional variations in refractive index, but does not include absorption. The treatment is rather different, but it may be shown that when Booker and Gordon's analysis is applied to ionospheric scattering, it gives the same result for the total power scattered as is obtained from the present analysis when a similar autocorrelation function for the ionization irregularities is assumed and absorption is neglected.

In the present case, an upgoing plane wave of angular frequency ω (wavelength λ) is considered to pass normally through a region of thickness t in which random variations of ionization density occur, with standard deviation n_0. The "scale" of the irregularities, a, defined in the Appendix, is assumed to be large compared with λ. The collisional frequency in the region is assumed to have the constant value ν. The wave after passing through the region contains lateral irregularities in phase and amplitude, which are shown to be equivalent to irregularities in phase only, and the r.m.s. value of the effective phase irregularity is given by

$$\phi_g = \frac{\pi^{1/4} n_0 e^2}{2mc\epsilon_0}\sqrt{\left(\frac{at}{\omega^2+\nu^2}\right)}$$

After a second, downward, passage through the irregular region, the wave as received at the ground is shown to consist of a specularly reflected component, or "signal," along with an angular distribution of scattered waves, or "noise." When ϕ_g is small (less than about 0·5rad), it is shown that the signal/noise ratio, i.e. the square root of the ratio of power in the signal and noise, is approximately $1/\sqrt{(2)}\phi_g$. Now, this ratio has been measured, and typical values of about 2·5 have been found for a wavelength of 60m (frequency 5 Mc/s), under quiet ionospheric conditions.[2] This leads to a value of ϕ_g of 0·3rad, corresponding to an r.m.s. fluctuation in phase path, for passage through the whole irregular region, of 3m.

With an angular distribution of waves as described, phase variations between the various components received at a fixed point on the ground will be caused either by random changes or by a systematic horizontal drift of the ionospheric irregularities, and these will give rise to the observed fading and directional fluctuations of the signal. The mean direction observed will, of course, correspond to that of the centre of the angular distribution of incident energy. It is shown in the Appendix that the r.m.s. angular deviation observed at the ground using a phase-comparison direction-finder whose aperture is not too large would be $\sigma = \lambda\phi_g/\sqrt{(2)}\pi a$. Hence if the above values of λ and ϕ_g are substituted, taking the value of σ (for $\lambda = 60$m) as 0·4° from Table 1, a value of 500m is obtained for a. This value is of the same order of magnitude as deduced by Briggs and Phillips[9] from fading experiments.

It is seen from the above equations that provided ν^2 is small compared with ω^2, which may be assumed to hold for the frequencies used, at heights greater than 75km, ϕ_g is proportional to λ, and hence σ is proportional to λ^2, as has been found experimentally. In this case, a value of $5 \times 10^{10}\,\mathrm{m}^{-5/2}$ is obtained for the product $n_0 t^{1/2}$. For instance, if the irregularities extend over a thickness of 10km, the r.m.s. fluctuation n_0 of the number of electrons per unit volume would be 500 per cubic centimetre, which is of the order of 5% of the daytime electron density in the lower part of the E region. In view of the various approximations introduced in the calculations, the numerical values of the parameters quoted can only be regarded as giving the orders of magnitude which would serve to explain the experimental results.

The mechanism described would provide an explanation of the quiet daytime fluctuations. It is probable that an additional agency, namely irregularities in the F layer, would be necessary to account for the larger fluctuations observed on F echoes during an ionospheric storm or under spread-F conditions. For instance, Dieminger[10] has given evidence that irregularities in both the E and F layers contribute towards the spread-F phenomenon. Also, large fluctuations of stellar radio signals are well correlated with spread-F, and Hewish[11] has deduced that the irregularities responsible for these fluctuations are at a height of about 400km. It may be noted that these observations also

showed angular fluctuations proportional to λ^2, so that a similar mechanism to that discussed above may well be responsible.

(3.2) Comparison of Vertical- and Oblique-Incidence Results

The oblique-incidence bearings show larger values of σ than the angular deviations at vertical incidence on the same frequency. An increased value would be expected owing to three separate factors: first, there would be an increase by a factor equal to the secant of the angle of elevation, since the bearing spread is measured in the horizontal plane instead of in a plane inclined at this angle to the horizontal; secondly, the angular fluctuations have been shown to be proportional to the square root of the path length through the irregularities, which introduces a factor equal to the square root of the secant of the angle of incidence at the irregular region. The product of these two factors is $1 \cdot 6$ for an F echo and $1 \cdot 8$ for an E echo on the 700 km oblique path. The third factor giving rise to increased bearing spreads at oblique incidence is the existence of random site errors, mentioned in Section 1. From investigations of the magnitude of such errors on the direction-finding system used[3] it appears that they could well account for the residual difference between the F-layer values of σ at vertical and oblique incidence.

The oblique-incidence E-layer results do not appear to be capable of such a simple explanation, since they do not show any dependence on frequency. The frequency range concerned, 5–10 Mc/s, corresponds as regards height of reflection to vertical-incidence frequencies of about $1 \cdot 7$–$3 \cdot 3$ Mc/s, in which range the various strata and "ledges" in the E region are most commonly found.[12] This may account qualitatively for these anomalous results, along with the scattered values obtained in this frequency range at vertical incidence.

(3.3) Application to Practical Direction-Finding

The results described have shown that with undisturbed ionospheric conditions, the rapid fluctuations in the measured direction of arrival of a single component of a reflected wave are only of the order of 1° or less. This applies to frequencies above about 3 Mc/s, and greater values may be expected at lower frequencies. The measurements were made with a direction-finder of 100m aperture; in the case of oblique incidence the figures quoted are not solely a direct result of the coning of the incident radiation but include an effect due to re-radiation from objects around the site. For systems of narrower aperture some increase in the latter component of variance would be expected, but it is not possible on present evidence to assign a numerical estimate.

In the presence of other direction-finding errors, e.g. lateral deviation, and more particularly, as in the usual case of continuous-wave working, wave-interference errors between components propagated by different modes, these fluctuations would make only a small contribution to the total errors observed. For instance, multi-mode wave-interference errors produced on an Adcock direction-finder have a variance commonly of the order of 5 degrees squared.[13] During ionospheric-storm or spread-F conditions, the fluctuations considered are much larger and may then be significant, but generally the effects of multi-mode interference will be predominant on a small-aperture system such as the Adcock.

When the angular spread of the downcoming wave does constitute a major source of bearing error, it has been shown that the rate of fluctuation is sufficiently rapid for effective smoothing to be carried out by taking bearings at 5-sec intervals. In Section 6.5, the correlation between bearings at separated direction-finders is considered, and eqn. (48) shows that this correlation falls to zero at a separation, transverse to the direction of propagation, of $a/\sqrt{2}$, i.e. about 400m. Thus two direction-finders at this separation would provide effective space-averaging of the bearings, in a similar way to the time-averaging obtained by taking observations at intervals of a few seconds at a single station. This distance, which is independent of frequency within the limits of application of the theory, would be reduced by the presence of random site errors which are uncorrelated at the two direction-finders.

(4) ACKNOWLEDGMENTS

The work described has extended over several years, and the author wishes to acknowledge the assistance of a number of his colleagues during this time, in carrying out the experiments and analysing the results. The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(5) REFERENCES

(1) Bramley, E. N., and Ross, W.: "Measurements of the Direction of Arrival of Short Radio Waves reflected at the Ionosphere," *Proceedings of the Royal Society*, A, 1951, **207**, p. 251.

(2) Bramley, E. N.: "Diversity Effects in Spaced-Aerial Reception of Ionospheric Waves," *Proceedings I.E.E.*, Paper No. 1062 R, January, 1951 (**98**, Part III, p. 19).

(3) Ross, W., Bramley, E. N., and Ashwell, G. E.: "A Phase-Comparison Method of Measuring the Direction of Arrival of Ionospheric Radio Waves," *ibid.*, Paper No. 1134 R, July, 1951 (**98**, Part III, p. 294).

(4) Piggott, W. R.: "The Reflection and Absorption of Radio Waves in the Ionosphere," *ibid.*, Paper No. 1434 R, March, 1953 (**100**, Part III, p. 61).

(5) Bain, W. C.: "On the Rapidity of Fluctuations in Continuous-Wave Radio Bearings at High Frequencies (see page 541).

(6) Bailey, D. K., *et al.*: "A New Kind of Radio Propagation at Very High Frequencies observable over Long Distances," *Physical Review*, 1952, **86**, p. 141.

(7) Jones, R. E., Millman, G. H., and Nertney, R. J.: "The Heights of Ionospheric Winds as measured at Long Radio Wavelengths," *Journal of Atmospheric and Terrestrial Physics*, 1953, **3**, p. 79.

(8) Booker, H. G., and Gordon, W. E.: "A Theory of Radio Scattering in the Troposphere," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 401.

(9) Briggs, B. H., and Phillips, G. J.: "A Study of the Horizontal Irregularities of the Ionosphere," *Proceedings of the Physical Society*, B, 1950, **63**, p. 907.

(10) Dieminger, W.: "The Scattering of Radio Waves," *ibid.*, 1951, **64**, p. 142.

(11) Hewish, A.: "The Diffraction of Galactic Radio Waves as a Method of Investigating the Irregular Structure of the Ionosphere," *Proceedings of the Royal Society*, A, 1952, **214**, p. 494.

(12) Whale, H. A.: "Fine Structure of the Ionospheric Region E," *Journal of Atmospheric and Terrestrial Physics*, 1951, **1**, p. 233.

(13) Smith, S. B., and Hopkins, H. G.: "High Frequency Direction Finding: Comparison of C.W. and Pulsed Transmissions," *Wireless Engineer*, 1954, **31**, p. 11.

(14) Booker, H. G., Ratcliffe, J. A., and Shinn, D. H.: "Diffraction from an Irregular Screen with Applications to Ionospheric Problems," *Philosophical Transactions of the Royal Society*, A, 1950, **242**, p. 579.

(6) APPENDIX: THEORY OF RAPID DIRECTIONAL FLUCTUATIONS PRODUCED BY IONOSPHERIC IRREGULARITIES

(6.1) Assumed Model of Irregularities

Consider a plane wave (angular frequency ω, free-space wavelength λ) passing normally through a horizontal slab of the ionosphere, of total thickness t, within which the electron density N has the same average value N_0 at all levels. Small variations of N, with a normal distribution, are assumed to occur in the vertical direction z and also horizontally except in one direction, y. The surfaces of constant density are thus cylinders along the y-axis, and normal to this axis (i.e. in the xz plane) the variations are statistically similar in all directions. Let the auto-correlation function of N at points separated by a distance r in the xz plane be of the form $\exp(-r^2/a^2)$, i.e. the correlation between the values of N at the points (x_1, y_1, z_1) and (x_2, y_2, z_2) is $\exp\left\{-[(x_1 - x_2)^2 + (z_1 - z_2)^2]/a^2\right\}$. It is assumed that the "scale" of the irregularities, a, is large compared with λ, so that only small changes in refractive index occur within a wavelength. It is also assumed that the collisional frequency ν is constant throughout the region considered.

(6.2) The Field of the Wave emerging from the Irregular Region

In passing through the total thickness, t, the incident wave will suffer a certain phase shift and a certain absorption, both of which will be random functions of the co-ordinate x. Let $\phi(x)$ and $\kappa(x)$ respectively be the excess phase shift and absorption, after traversing the thickness t, compared with the values which would exist in the absence of the irregularities. Owing to variations in ϕ and κ as x is varied, the emergent wave may be represented as an angular distribution of energy, which may be specified in terms of the statistical properties of ϕ and κ. These will now be determined.

Consider the phase shift and absorption produced over an element of thickness dz, where the electron density is $N = N_0 + dN$. The phase shift introduced is $(2\pi/\lambda)\mu dz$, where μ is the refractive index. Using M.K.S. units, and neglecting the effect of the earth's magnetic field, μ is given by

$$\mu^2 = 1 - \frac{Ne^2}{m\epsilon_0(\omega^2 + \nu^2)} \quad . \quad . \quad . \quad . \quad (1)$$

where e and m are the electronic charge and mass, and ϵ_0 is the permittivity of free space. A small change dN in the electronic density thus produces a change in μ given by

$$d\mu = \frac{e^2 dN}{2\mu m\epsilon_0(\omega^2 + \nu^2)}$$

and the additional phase shift, relative to the undisturbed value, will be

$$d\phi = \frac{\pi e^2 dN dz}{\mu\lambda m\epsilon_0(\omega^2 + \nu^2)} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If the critical frequency of the region where the irregularities are situated is well below the working frequency, μ will not differ much from unity. Also, if ω_0 is 2π times this critical frequency,

$$\frac{e^2}{m\epsilon_0} = \frac{\omega_0^2 + \nu^2}{N_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and eqn. (2) may be written

$$d\phi = \frac{\pi}{\lambda}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{dN}{N_0}dz \quad . \quad . \quad . \quad . \quad (4)$$

Similarly, the absorption suffered by the wave in the element dz is kdz where k is the absorption coefficient given by

$$k = \frac{\nu}{2c\mu}\frac{Ne^2}{m\epsilon_0(\omega^2 + \nu^2)} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where c is the speed of light in free space; thus the extra absorption introduced by the increase dN over the distance dz is

$$d\kappa = \frac{\nu}{2c}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{dN}{N_0}dz \quad . \quad . \quad . \quad . \quad (6)$$

The changes in phase and absorption thus bear a constant ratio to each other, given by

$$\frac{d\kappa}{d\phi} = \frac{\nu\lambda}{2\pi c} = \frac{\nu}{\omega} = g \text{ (say)} \quad . \quad . \quad . \quad . \quad (7)$$

and since this applies to every element dz it will also hold for the values ϕ and κ appropriate to the whole distance t.

i.e.
$$\kappa = g\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The statistical properties of ϕ and κ can now be found from those of N. It is assumed that N has standard deviation n_0, and its auto-correlation function in the z-direction is $\rho(z) = \exp(-z^2/a^2)$.

From eqn. (4) the mean-square phase-shift per unit distance of z is

$$\overline{\left(\frac{d\phi}{dz}\right)^2} = \left(\frac{\pi}{\lambda}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{n_0}{N_0}\right)^2$$

Now if any randomly varying function has variance v and auto-correlation coefficient $\rho(s)$ for the interval s, then the variance of its average value over an interval t may be shown to be

$$2v\int_0^1 (1 - u)\rho(ut)du$$

Hence the variance ϕ_0^2 of the total phase shift ϕ over the length t, when ϕ is taken over all values of x, is given by

$$\begin{aligned}\phi_0^2 &= t^2\overline{\left(\frac{d\phi}{dz}\right)^2}2\int_0^1 (1 - u)\rho(ut)du \\ &= 2t^2\left(\frac{\pi}{\lambda}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{n_0}{N_0}\right)^2\int_0^1 (1 - u)\varepsilon^{-u^2t^2/a^2}du \\ &= t^2\left(\frac{\pi}{\lambda}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{n_0}{N_0}\right)^2\left[\frac{\sqrt{(\pi)}a}{t}\operatorname{erf}\left(\frac{t}{a}\right) - \frac{a^2}{t^2}(1 - \varepsilon^{-t^2/a^2})\right]\end{aligned} \quad . \quad . \quad . \quad . \quad (9)$$

If the total thickness t of the irregular region is assumed to be large compared with a, the scale of the irregularities, this reduces to

$$\phi_0^2 = \sqrt{(\pi)}at\left(\frac{\pi}{\lambda}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{n_0}{N_0}\right)^2 \quad . \quad . \quad . \quad (10)$$

and the r.m.s. value ϕ_0 will then be

$$\phi_0 = \frac{\pi^{5/4}}{\lambda}\frac{\omega_0^2 + \nu^2}{\omega^2 + \nu^2}\frac{n_0}{N_0}(at)^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (11)$$

Also, since ϕ (the excess phase-shift) is a linear function of the elemental phase-shifts along the path, which are normally distributed with zero mean, ϕ itself is also normally distributed with zero mean. Similarly, the total excess absorption due to the irregularities is distributed in the same way, and has an r.m.s. value given by

$$\kappa_0 = \frac{\nu}{\omega}\phi_0 = g\phi_0 \quad . \quad . \quad . \quad . \quad . \quad (12)$$

To complete the statistical specification of the field after traversing the distance t, from which the resultant angular distribution is to be found, it is necessary to know the auto-correlation function for ϕ (or κ) in the x-direction. Thus if $\phi(x)$ and $\phi(x + \xi)$ are the values of ϕ at two points separated by a distance ξ in the x-direction, this correlation function is

given by $\rho(\xi) = \overline{\phi(x)\phi(x+\xi)}/\phi_0^2$, the mean being taken over all values of x.

Now

$$\phi(x) = \int_0^t \left(\frac{d\phi}{dz}\right)_x dz, \text{ and } \phi(x+\xi) = \int_0^t \left(\frac{d\phi}{dz}\right)_{x+\xi} dz$$

Hence

$$\rho(\xi) = \frac{1}{\phi_0^2}\int_0^t\int_0^t \overline{\left(\frac{d\phi}{dz}\right)_x\left(\frac{d\phi}{dz}\right)_{x+\xi}} dz_1 dz_2 \quad . \quad (13)$$

Since the auto-correlation function for N, and hence also for $d\phi/dz$, is of the form $\exp(-r^2/a^2)$, this may be written

$$\rho(\xi) = \frac{1}{\phi_0^2}\left(\frac{\pi}{\lambda}\frac{\omega_0^2+\nu^2}{\omega^2+\nu^2}\frac{n_0}{N_0}\right)^2 \times$$
$$\exp(-\xi^2/a^2)\int_0^t\int_0^t \exp[-(z_1-z_2)^2/a^2]dz_1dz_2 \quad . \quad (14)$$

Evaluation of the double integral gives

$$\int_0^t\int_0^t \exp[-(z_1-z_2)^2/a^2]dz_1dz_2$$
$$= a\{\sqrt{(\pi)}t \operatorname{erf}(t/a) - a[1-\exp(-t^2/a^2)]\} \cdot (15)$$

and hence, substituting for ϕ_0^2 from eqn. (9),

$$\rho(\xi) = \varepsilon^{-\xi^2/a^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

i.e. the auto-correlation function for ϕ (or κ) in the x-direction, is of the same simple form as that for the variations in electronic density.

The amplitude $E(x)$ of the emergent wave is given by

$$E(x) = E_0\varepsilon^{-\kappa(x)} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where E_0 is the amplitude which would be obtained in the absence of irregularities. Now κ is normally distributed with zero mean and standard deviation κ_0; i.e. its probability distribution is

$$p(\kappa) = \frac{1}{\sqrt{(2\pi)}\kappa_0}\varepsilon^{-\kappa^2/2\kappa_0^2} \quad . \quad . \quad . \quad . \quad (18)$$

The mean-square value of E is then given by

$$\overline{E^2} = \int_{-\infty}^{+\infty} E^2 p(\kappa)d\kappa$$
$$= \frac{E_0^2}{\sqrt{(2\pi)}\kappa_0}\int_{-\infty}^{+\infty}\exp[-(2\kappa+\kappa^2/2\kappa_0^2)]d\kappa$$
$$= E_0^2\varepsilon^{2\kappa_0^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Eqns. (8), (11), (16) and (19) completely specify the statistical properties of the wave field immediately after passing through the thickness t.

(6.3) The Angular Distribution of the Energy after Passing through the Irregular Region

The field at the upper boundary of the irregular region may be written

$$E(x) = E(x)\varepsilon^{j\phi(x)} = E_0\varepsilon^{-\kappa(x)+j\phi(x)}$$
$$= E_0\varepsilon^{-(g-j)\phi(x)} \text{ [from eqn. (8)]} \quad . \quad . \quad . \quad (20)$$

The angular distribution is to be found from the auto-correlation function for this complex field, defined by

$$\rho_E(\xi) = \frac{\overline{E^*(x)E(x+\xi)}}{\overline{E(x)E^*(x)}} \quad . \quad . \quad . \quad . \quad (21)$$

where the asterisk denotes the complex conjugate and the bar indicates an average taken over all values of x. Thus

$$\rho_E(\xi) = (E_0^2/\overline{E^2})\overline{\exp[-(g+j)\phi(x)]\exp[-(g-j)\phi(x+\xi)]} \quad . \quad . \quad . \quad . \quad (22)$$

Hence from eqn. (19),

$$\rho_E(\xi)\exp(2\kappa_0^2) =$$
$$\overline{\exp\{-g[\phi(x)+\phi(x+\xi)]+j[\phi(x+\xi)-\phi(x)]\}}$$
$$= \overline{\exp\{-g[\phi(x)+\phi(x+\xi)]\}\cos[\phi(x+\xi)-\phi(x)]} \quad . \quad . \quad . \quad . \quad (23)$$

Now $\phi(x)$ and $\phi(x+\xi)$ (subsequently written as ϕ_1 and ϕ_2 for brevity) are each normally distributed about zero mean with standard deviation ϕ_0 and with correlation coefficient $\rho(\xi) = \varepsilon^{-\xi^2/a^2}$. Their joint probability distribution is therefore

$$p(\phi_1,\phi_2) =$$
$$\frac{1}{2\pi\phi_0^2\sqrt{(1-\rho^2)}}\exp[-(\phi_1^2-2\rho\phi_1\phi_2+\phi_2^2)/2\phi_0^2(1-\rho^2)] \quad . \quad . \quad . \quad . \quad (24)$$

Now the mean value of a function of these variables, say $q(\phi_1,\phi_2)$, is given by

$$\bar{q} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} q(\phi_1,\phi_2)p(\phi_1,\phi_2)d\phi_1 d\phi_2 \quad . \quad . \quad (25)$$

Hence

$$\rho_E(\xi)\exp(2\kappa_0^2) = \frac{1}{2\pi\phi_0^2\sqrt{(1-\rho^2)}} \times$$
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\varepsilon^{-g(\phi_1+\phi_2)}\cos(\phi_1-\phi_2)\varepsilon^{-\frac{\phi_1^2-2\rho\phi_1\phi_2+\phi_2^2}{2\phi_0^2(1-\rho^2)}}d\phi_1d\phi_2$$
$$= \exp\{-\phi_0^2[(1-\rho)-g^2(1+\rho)]\} \quad . \quad . \quad . \quad (26)$$

Therefore

$$\rho_E(\xi) = \exp[-\phi_0^2(1+g^2)(1-\rho)] \quad . \quad . \quad (27)$$

Thus, in the absence of absorption, when the layer behaves as a phase-changing screen only, the result is

$$\rho_E(\xi) = \exp[-\phi_0^2(1-\rho)] \quad . \quad . \quad . \quad . \quad (28)$$

and the effect of absorption is simply to increase the effective mean-square phase variation by the factor $(1+g^2) = (1+\nu^2/\omega^2)$. It now follows from a relation [their eqn. (15)] given by Booker, Ratcliffe and Shinn,[14] between the auto-correlation function $\rho_E(\xi)$ and the angular power spectrum, that in the case of a narrow angular distribution $P(\theta)$, this is given by

$$P(\theta) = \frac{P_0}{\lambda}\int_{-\infty}^{+\infty}\rho_E(\xi)\varepsilon^{-j\frac{2\pi\theta}{\lambda}\xi}d\xi \quad . \quad . \quad . \quad (29)$$

where $P(\theta)$ is the power density per unit angle in the direction θ to the normal to the screen, and $P_0 = \int_{-\infty}^{+\infty}P(\theta)d\theta$ is the total power in the wave. From eqns. (16), (27) and (29) it is found that

$$P(\theta) = P_0\varepsilon^{-\phi_g^2}\delta(\theta) + \frac{\sqrt{(\pi)}a}{\lambda}P_0\varepsilon^{-\phi_g^2}\sum_{n=1}^{\infty}\frac{\phi_g^{2n}}{n!}\frac{\varepsilon^{-\frac{\pi^2a^2\theta^2}{n\lambda^2}}}{\sqrt{n}} \quad (30)$$
$$= P_0\varepsilon^{-\phi_g^2}\delta(\theta) + P_1(\theta) \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where

$$\phi_g^2 = \phi_0^2(1+g^2) = \frac{\pi^{1/2}}{4c^2}\left(\frac{n_0e^2}{m\epsilon_0}\right)^2\frac{at}{\omega^2+\nu^2}$$
$$\text{from eqns. (3), (11) and (12)} \quad . \quad (32)$$

The first term on the right-hand side of eqn. (30) is a Dirac delta function at $\theta = 0$, which means a pure "signal" as contrasted with the remainder of the distribution, which may be regarded as "noise." The total energy in the "signal" is $P_0\varepsilon^{-\phi_g^2}$, and the "signal/noise ratio," i.e. the square root of the ratio of the signal and noise powers, is

$$b_1 = [\varepsilon^{-\phi_g^2}/(1 - \varepsilon^{-\phi_g^2})]^{1/2} \quad . \quad . \quad . \quad . \quad (33)$$

When ϕ_g is small (less than about 0·5 rad) this reduces to

$$b_1 = 1/\phi_g \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad 34)$$

The second term in the distribution given by eqn. (30) represents a continuous angular distribution, which has r.m.s. value θ_0 given by

$$\theta_0^2 = \int_{-\infty}^{+\infty} \theta^2 P_1(\theta)d\theta \bigg/ \int_{-\infty}^{+\infty} P_1(\theta)d\theta$$

$$= \frac{\lambda^2\phi_g^2}{2\pi^2a^2(1 - \varepsilon^{-\phi_g^2})} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

and again, when ϕ_g is small, this reduces to

$$\theta_0 = \lambda/\sqrt{(2)}\pi a \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

It may be noted that for the case of small ϕ_g a useful approximation for $P_1(\theta)$ is obtained by taking the first term only of the summation in eqn. (30). This gives

$$P_1(\theta) = \frac{a\sqrt{(\pi)}}{\lambda}P_0\varepsilon^{-\phi_g^2}\phi_g^2\varepsilon^{-\pi^2a^2\theta^2/\lambda^2}$$

which represents a simple Gaussian distribution of the type previously considered.[2] It has the correct r.m.s. value as given by eqn. (36), and the total energy in the distribution is $P_0\phi_g^2\varepsilon^{-\phi_g^2}$ which is a good approximation to the value of $P_0(1 - \varepsilon^{-\phi_g^2})$ obtained by integrating the full expression for $P_1(\theta)$.

(6.4) Angular Fluctuations observed at the Ground

Section 6.3 relates to the single passage of a wave through the irregular region. If this occurs on the upward journey of the wave, which is subsequently reflected specularly, the distribution will be further modified by a second traversal of the region on the downward journey. In the case considered, where ϕ_0 is small and b_1 large compared with unity, this second transit may be allowed for simply by considering the total thickness of the irregular region to be doubled. It is then easily seen that the signal/noise ratio of the wave as received at the ground will be given approximately by

$$b = b_1/\sqrt{(2)} = 1/\sqrt{(2)}\phi_g \quad . \quad . \quad . \quad . \quad . \quad (37)$$

while the angular distribution of the noise is substantially the same as given by eqn. (36).

Now the correlation function R for two points a distance d apart, in the x direction on the ground (e.g. the two aerials of a phase-comparison direction-finder), in the presence of the angular distribution $P_1(\theta)$, is defined by [Reference 2, eqn. (27)]

$$R = \int_{-\infty}^{+\infty} P_1(\theta)\cos\left(\frac{2\pi d}{\lambda}\theta\right)d\theta \bigg/ \int_{-\infty}^{+\infty} P_1(\theta)d\theta$$

which in the present case gives

$$R = \frac{\varepsilon^{-\phi_g^2}}{1 - \varepsilon^{-\phi_g^2}}[\exp(\phi_g^2\varepsilon^{-d^2/a^2}) - 1] \quad . \quad . \quad . \quad (38)$$

reducing to

$$R = \varepsilon^{-d^2/a^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

in the case of small ϕ_g.

The r.m.s. phase difference Φ_0 observed at the two points distance d apart is then given by (Reference 2, Section 6)

$$\Phi_0 = \frac{1}{b}(1 - R)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

and hence the r.m.s. fluctuation in direction of arrival is

$$\sigma = \lambda\Phi_0/2\pi d$$

$$= \lambda(1 - R)^{1/2}/2\pi db \quad . \quad . \quad . \quad . \quad (41)$$

Substituting the values of R and b from eqns. (39) and (37),

$$\sigma = \frac{\lambda\phi_g}{\sqrt{(2)}\pi d}(1 - \varepsilon^{-d^2/a^2})^{1/2} \quad . \quad . \quad . \quad (42)$$

which simplifies, if $d^2/a^2 \ll 1$, to

$$\sigma = \frac{\lambda\phi_g}{\sqrt{(2)}\pi a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Substituting for ϕ_g from eqn. (32), this gives

$$\sigma = \frac{\pi^{1/4}e^2n_0}{m\epsilon_0\omega}\left[\frac{t}{2a(\omega^2 + \nu^2)}\right]^{1/2} \quad . \quad . \quad . \quad (44)$$

This gives the r.m.s. fluctuation in direction of arrival in terms of the various parameters specifying the irregular properties of the medium.

(6.5) Correlation between Bearings at Separated Direction-Finders

Consider two direction-finders, each operating on the phase-comparison principle, with aerial spacing d, separated by a distance ξ in the x-direction. At the ground, let $\rho_1(\xi)$ be the auto-correlation function for phase along the x-direction. It has been shown[2] that for the case of large signal/noise ratio this is the same as R, the correlation function for the angular distribution. In the present case this is of the form $\varepsilon^{-\xi^2/a^2}$ [eqn. (39)], which is the same as the auto-correlation function $\rho(\xi)$ for phase at the boundary of the irregular region itself. Thus

$$\rho_1(\xi) = \rho(\xi) = \varepsilon^{-\xi^2/a^2} \quad . \quad . \quad . \quad . \quad (45)$$

The simultaneous bearings obtained at the direction-finders (separated by ξ) are each determined by the phase difference between the receiving aerials (spacing d for each instrument). The correlation coefficient for these phase differences is easily shown to be

$$\rho(d, \xi) = \frac{2\rho(\xi) - \rho(\xi + d) - \rho(\xi - d)}{2[1 - \rho(d)]} \quad . \quad . \quad (46)$$

$$= \frac{\varepsilon^{-\xi^2/a^2}[1 - \varepsilon^{-d^2/a^2}\cosh(2\xi d/a^2)]}{1 - \varepsilon^{-d^2/a^2}} \quad . \quad (47)$$

This reduces to

$$\rho(d, \xi) = (1 - 2\xi^2/a^2)\varepsilon^{-\xi^2/a^2} \quad . \quad . \quad . \quad (48)$$

when d/a is small, i.e. for narrow-aperture direction-finders. In this case the correlation coefficient for the bearings is seen to fall to zero at a separation of $a/\sqrt{(2)}$, which is of the order of 400 m with the value of a deduced from the experimental results. By substitution of values into eqn. (47) it is seen that this conclusion is not substantially affected when d is as much as 100 m, as in the present experiments.

[The discussion on the above paper will be found on page 550.]

621.396.93 : 621.396.11

Paper No. 1715 R
Oct. 1954

ON THE RAPIDITY OF FLUCTUATIONS IN CONTINUOUS-WAVE RADIO BEARINGS AT HIGH FREQUENCIES

By W. C. BAIN, M.A., B.Sc., Ph.D.

(*The paper was first received* 11*th March, and in revised form* 20*th May*, 1954. *It was published in October*, 1954, *and was read before the* RADIO SECTION 2*nd March*, 1955.)

SUMMARY

Bearing observations were made with an Adcock direction-finder on distant high-frequency transmitters at a rate of five per second. The auto-correlation function of the observations was calculated; the analytical formula found most useful as its representation was $\varepsilon^{-\tau/\tau_0}$, where the mean value of τ_0 was $0 \cdot 75$ sec with a standard deviation of $0 \cdot 51$ sec. Expressions are given for the reduction in the variance of bearing error to be obtained by time-averaging.

(1) INTRODUCTION

In high-frequency direction-finding with an Adcock system, the observed bearing of a remote station is usually found to fluctuate owing to changes in the ionosphere. Some of the fluctuations—those caused by lateral deviation in the ionosphere—have quasi-periods of the order of 20 min, and have been investigated in some detail by Bramley and Ross.[1] However, an examination of the bearing display given on a cathode-ray direction-finder, for instance, by a continuous-wave station will reveal in general the presence of far more rapid fluctuations. These are due to interference between rays which have travelled to the direction-finder along different paths, and are often of a much larger magnitude than those due to lateral deviation. Their effect upon the accuracy with which the bearing of a station is known might be expected to be reduced considerably by averaging over a comparatively short period of time, say 10 or 20 sec.

A quantitative measure of the extent of this improvement is required, for example, in the theory developed in connection with obtaining a position fix by radio direction-finding. In this theory an area can be determined within which there is a given probability of the transmitter being found, provided that a known variance can be assigned to the bearing given by each direction-finding station. If time-averaging is employed to obtain such a bearing, a reduction can be made in the appropriate component of this variance, with a corresponding diminution of the search area. The extent of this reduction will have to be based on information about the rapidity of the bearing variations.

An investigation was therefore made at the Radio Research Station on the properties of these rapid fluctuations. The object was not to make an exhaustive study of the problem, but to collect sufficient information to make possible a reasonably good estimate of the reduction in the error variance secured by averaging over a given period of time.

(2) EXPERIMENTAL TECHNIQUE

The direction-finding system on which the experiments were carried out was the U-Adcock with buried feeders installed at the Radio Research Station, Slough. The receiver used was a cathode-ray direction-finding set of the type FHB as supplied to the Admiralty.[2] It was decided to record the bearings photographically, and a simple camera was therefore constructed which was capable of taking pictures of the cathode-ray-tube screen in the receiver at a rate of up to five per second on 10 cm film. The method of operation was to tune in to the desired station, to perform the usual lining-up procedure on the twin-channel receiver, and then to take about 100 pictures in succession of the varying display on the tube. Broadcast stations were used to avoid missed pictures due to gaps in transmission, such as those between Morse characters; details concerning the transmitters observed are given in Table 1.

Table 1

Location of transmitter	Frequency	Bearing from Slough	Distance
	Mc/s	deg	km
Schwarzenburg ..	6·16	129	780
Leipzig	9·73	86	880
Delhi	17·7	80	6 740
Zanzibar	19·1	136	7 560

The indicated bearings were read off direct from the developed film, and plotted out against time as in the examples shown in Fig. 1.

(3) THE ANALYSIS OF THE OBSERVATIONAL DATA

The reduction in the variance of the mean obtained by averaging can be calculated from the auto-correlation function of the observations, and so for each group of bearing measurements this function was computed. Now the value of the auto-correlation function at a time equal to one interval between successive pictures provides a check on whether the camera is being operated quickly enough to follow the fluctuations adequately. In these experiments the time interval was reduced until the majority of such values exceeded $0 \cdot 5$. This involved operating the camera near its maximum rate of five pictures per second.

Some groups of bearing observations and the auto-correlation functions deduced from them are shown in Fig. 1. Even in these examples there is considerable variety in the form of the function, and still other types are to be found among those not illustrated here. Now to calculate the variance reduction in a general case, an analytical expression is required to cover all possible forms of auto-correlation function. It is clear, however, that such a formula would have to be rather complicated and to contain numerous parameters in order to give a close fit to the observed functions. A simple expression with one parameter was therefore chosen for the theoretical auto-correlation function, which was to be fitted as well as possible to the observed values and would represent their behaviour on the average, if not in detail. This was the exponential form $R(\tau) = \varepsilon^{-\tau/\tau_0}$, which has been frequently used in the theory of random functions.

A value has to be found for the parameter τ_0 appropriate to each period of observations. The following standard procedure was adopted to evaluate τ_0 in each case. Suppose that at a time interval τ the observed correlation is found to be $R(\tau)$. The

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

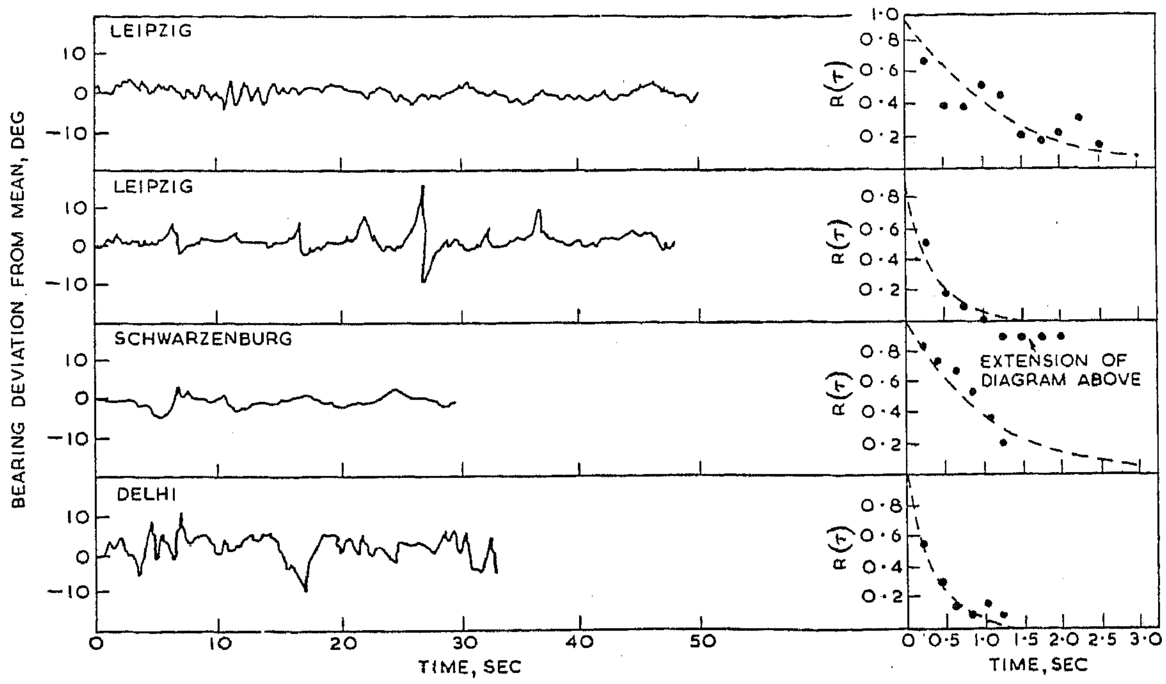


Fig. 1.—Bearing plots and corresponding auto-correlation functions, $R(\tau)$.

● ● ● Deduced from observations.

— — — Theoretical curve $\varepsilon^{-\tau/\tau_0}$.

The values of τ_0 deduced from the above plots are (reading from top to bottom of the Figure) 1·15, 0·30, 0·99 and 0·33 sec.

value of $\log_\varepsilon R$ is obtained and plotted against τ; this is repeated for all points up to but not including the first point for which $R(\tau)$ is less than 0·1. All these points will lie on a straight line if $R(\tau)$ has an exponential form. In general they do not, and the "best" straight line has to be drawn through them. The best line has been taken here to be the line joining the origin in the τ, $\log_\varepsilon R$ plane to the centroid of the remaining points. A small change in a value for $\log_\varepsilon R$ at a given τ has then the same effect on the slope of the line whatever the corresponding value of τ may be—a desirable property of the method, although its chief merit is the ease with which the position of the line can be calculated. Curves of $R(\tau)$ with the values of τ_0 calculated in this way have been included in Fig. 1 so that the degree of approximation involved in using the simple exponential form may be appreciated.

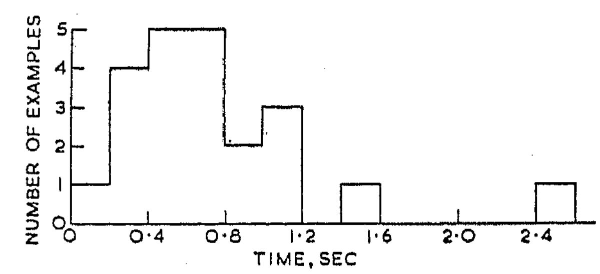


Fig. 2.—Number of examples of τ_0 falling within a given range of time.

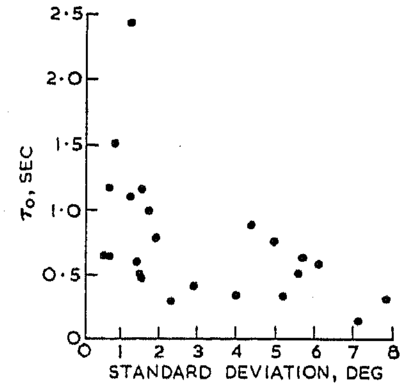


Fig. 3.—Relation between the parameter τ_0 and the standard deviation of a group of bearings.

(4) RESULTS

The groups of observations shown in Fig. 1 are typical of those obtained, but the data which have been collected are insufficient to determine whether a particular kind of fluctuation is characteristic of a particular station. It has therefore been thought unnecessary to give specimen records for each station observed.

The values obtained for τ_0 by the methods discussed in Section 3 are shown in the histogram of Fig. 2. This distribution has a mean of 0·75 sec and a standard deviation of 0·51 sec. Its range is from 0·13 to 2·43 sec and its mode is about 0·6 sec. It may be noted that Bramley[3] has found that, on pulse transmissions, bearings are uncorrelated for a time interval of 5 sec, but may be appreciably correlated at an interval of 1 or 2 sec. With the pulse transmissions the fluctuations are due to interference within a single ionospheric "ray," whereas in the present series of continuous-wave observations the fluctuations are caused mainly by interference between separate rays, at least in those cases where the bearing variance is high.

The standard deviation of the observations themselves varied considerably between the groups of readings used for each auto-correlation measurement. As it is possible that it is related to the value of τ_0 determined for each group, a scatter diagram for corresponding values of τ_0 and the standard deviation has been constructed, as shown in Fig. 3. It appears that the two parameters are not independent, there being no cases where a high standard deviation is associated with a high value of τ_0. The results were also examined for a connection between τ_0 and the distance of the transmitter, but no significant effect was found.

In practical direction-finding the reduction of small variances by time-averaging is of little importance—it is the reduction of the large variances which matters. Hence, in view of the relation mentioned above between τ_0 and the standard deviation, the mean value of τ_0 has been recalculated using only groups of observations in which the variance is more than 2 deg.^{2} The distribution of τ_0 has now a range 0·13 to 1·15 sec and a mean of 0·56 sec with standard deviation 0·29 sec; these values appear to be the most appropriate to use in estimating the improvement

due to time-averaging on occasions when appreciable errors consisting of rapid bearing fluctuations are present.

(5) THE REDUCTION OF ERROR VARIANCE BY TIME-AVERAGING

The reduction in variance obtained by averaging over N observations, spaced by a time interval τ_1 between each, when the auto-correlation function is $R(\tau) = \varepsilon^{-\tau/\tau_0}$, has been given, for example, by Burgess,[4] and is expressed by

$$\frac{\sigma_N^2}{\sigma^2} = \frac{N(1 - R_1^2) - 2R_1(1 - R_1^N)}{N^2(1 - R_1)^2}$$

where $R_1 = R(\tau_1) = \varepsilon^{-\tau_1/\tau_0}$.
σ^2 = Variance of one observation.
σ_N^2 = Variance of the mean of N observations.

If the observations are averaged continuously over a period of time T (i.e. $\tau_1 \to 0$), the variance of the mean σ_T^2 is given by:

$$\frac{\sigma_T^2}{\sigma^2} = 2\frac{\tau_0^2}{T^2}\left(\varepsilon^{-T/\tau_0} + \frac{T}{\tau_0} - 1\right)$$

Fig. 4 shows the reduction in variance to be expected as a result of time averaging with different numbers of observations

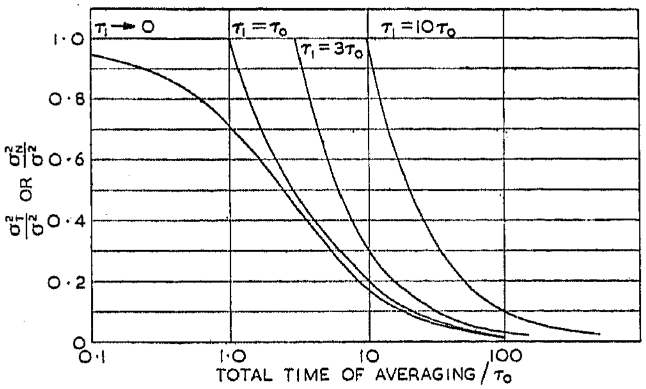


Fig. 4.—Reduction of variance by time-averaging.

per unit time; if a value of τ_0 of 0·56 sec is used here, the variance is reduced by a factor of 10 in 12 sec provided that the time interval between observations is kept below τ_0. If the value of 1·15 sec is used for τ_0, the graphs will serve as a guide to the most unfavourable results likely to be obtained in practice by the time-averaging of bearings.

(6) CONCLUSIONS

The auto-correlation functions calculated from the variations in bearing given by an Adcock direction-finder in the high-frequency band do not conform closely to any simple analytical form. They can be roughly represented by the exponential curve $R(\tau) = \varepsilon^{-\tau/\tau_0}$, where the mean value found for τ_0 is 0·75 sec with a standard deviation of 0·51 sec. It has been shown that when this function is used to calculate the reduction in variance of bearings with large fluctuations a value of τ_0 of 0·56 sec is more appropriate, since lower values of τ_0 tend to be associated with the large bearing variances.

(7) ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance rendered in these experiments by Mr. C. Jones, who constructed the camera mechanism used and operated it during most of the observational work. The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(8) REFERENCES

(1) BRAMLEY, E. N., and ROSS, W.: "Measurements of the Direction of Arrival of Short Radio Waves reflected at the Ionosphere," *Proceedings of the Royal Society*, A, 1951, **207**, p. 251.
(2) DE WALDEN, S., ROCKE, A. F. L., BARRETT, J. O. G., and PITTS, W. J.: "The Development of a High-Frequency Cathode-Ray Direction Finder for Naval Use," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 823.
(3) BRAMLEY, E. N.: "Some Aspects of the Rapid Directional Fluctuations of Short Radio Waves reflected at the Ionosphere" (see page 533).
(4) BURGESS, R. E.: "The Rectification and Observation of Signals in the Presence of Noise," *Philosophical Magazine*, 1951, **42**, p. 475.

[The discussion on the above paper will be found on page 550.]

621.396.11 : 621.317.34.029.5/.6

Paper No. 1716 R
Sept. 1954

SOME COMPARATIVE DIRECTIONAL MEASUREMENTS ON SHORT RADIO WAVES OVER DIFFERENT TRANSMISSION PATHS

By E. N. BRAMLEY, M.Sc.

(*The paper was first received* 11*th March, and in revised form* 20*th May*, 1954. *It was published in September*, 1954, *and was read before the* RADIO SECTION 2*nd March*, 1955.)

SUMMARY

Observations on first- and second-order F-layer echoes from a pulsed transmitter at frequencies of 5–8 Mc/s have been made at two direction-finders separated by 213 km. From a comparison of first-order bearings over the two paths it has been found that negligible correlation exists between similar components of the random ionospheric tilts at points separated by 106 km. However, relatively small average tilts over a period of an hour have been found to occur, and these are correlated over wide areas; they show a regular diurnal variation.

The mid-points of the two transmission paths lay on land and on sea respectively, and a comparison of the directional variations exhibited by the second-order echo showed that the land behaved as a rougher reflector than the sea. The resulting difference in the total variance of bearings on this echo for the two paths was, however, small, as the deviations were mainly caused by ionospheric tilts.

LIST OF PRINCIPAL SYMBOLS

h = Height of ionospheric layer.
D = Distance of transmission.
θ = Angle of tilt of ionospheric layer from the horizontal.
ψ = Angle between axis of layer tilt and the normal to vertical plane containing transmitter and receiver.
θ', θ'' = Values of θ at reflection points for second-order reflection.
ψ', ψ'' = Values of ψ at reflection points for second-order reflection.
α_1 = Bearing shift due to layer tilt on first-order reflection.
Δ_1 = Elevation shift due to layer tilt on first-order reflection.
α_2 = Bearing shift due to layer tilts on second-order reflection.
Δ_2 = Elevation shift due to layer tilts on second-order reflection.
δ_1, δ_2 = Angles of elevation of first- and second-order reflections.
x, y = Orthogonal displacements of ground reflection point on second-order reflection, due to layer tilts.
v_1 = Variance of bearing or elevation for first-order reflection.
v_2 = Variance of bearing or elevation for second-order reflection.
v_{1S} = Slow component of first-order variance.
v_{1R} = Rapid component of first-order variance.
v_{2S} = Slow component of second-order variance.
v_{2R} = Rapid component of second-order variance.
$k_S = v_{2S}/v_{1S}$.
$k_R = v_{2R}/v_{1R}$.
ρ = Correlation coefficient between ionospheric tilts at separated reflection points.
σ = Estimated standard deviation of ρ.

(1) INTRODUCTION

In recent years an extensive series of observations has been carried out on the direction of arrival of short radio waves reflected at the ionosphere. This work, which has been conducted using pulse-modulated transmissions at vertical and oblique incidence, has provided data on the limits of accuracy of high-frequency direction-finding, on first-order reflections, set by ionospheric irregularities. It has been shown that slow changes in the direction of arrival occur, with periods of the order of 20 min, and these are attributable to tilts of the surfaces of constant ionization density in the ionosphere.[1] Some experimental results have been obtained, using spaced receivers or transmitters, which indicate the lateral scale of magnitude of these tilts.[2] The slow changes are accompanied by more rapid fluctuations, due to wave interference within the ray and having periods of the order of a few seconds. The magnitude and causes of these are discussed in a separate paper.[3]

In the course of the experiments a few observations were made on second-order reflections on an oblique path,[1] and these were found to exhibit a marked increase in the magnitude of the rapid fluctuations as compared with a first-order echo. The difference was greater than that which would be expected solely as the result of two ionospheric reflections instead of one. It was suggested that this state of affairs was attributable to scattering of the radiation by the ground at the mid-point of the transmission path, which in this case was on land near Stockton-on-Tees. The question naturally arises whether such scattering would be appreciably different for a mid-point on the sea.

The experiments to be described were carried out with two objects in view:

(*a*) To obtain further data on the lateral scale of the ionospheric tilts mentioned above.
(*b*) To make a direct comparison of the effects on second-order reflections when the mid-points were on land and on sea respectively.

(2) EXPERIMENTAL ARRANGEMENTS AND METHOD

The experiments were carried out during the period from September, 1952, to March, 1953, and consisted of directional measurements at two different receiving sites on signals from a common transmitter. The transmitter was at Inverness, which is 704 km from the direction-finder at Winkfield, Berkshire, and 648 km from that at Hemsby, Norfolk. The mid-points of the two transmission paths were, respectively, on land near Kirkby Stephen, Westmorland, and on sea, about 20 km off the coast at Tynemouth. These points are 106 km apart in a direction 49° east of north. The geometry of the paths and the location of the various reflection points involved are shown in Fig. 1.

The two direction-finders were of the same phase-comparison type,[4] each with an aerial spacing of 100 m. The transmitter emitted pulse-modulated signals at various frequencies in the range 5–8 Mc/s, and the directional measurements were made over the same periods at the two receivers. These periods were mainly in the daytime, but a few extended for several hours after sunset. Attention was confined to the first- and second-order F-layer echoes, and in each case the direction of arrival was measured in terms of the bearing deviation, α, from the great-circle direction of the transmitter, and the angle of elevation, δ, above the horizontal. The individual observations, which were photographic, were normally made at regular

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

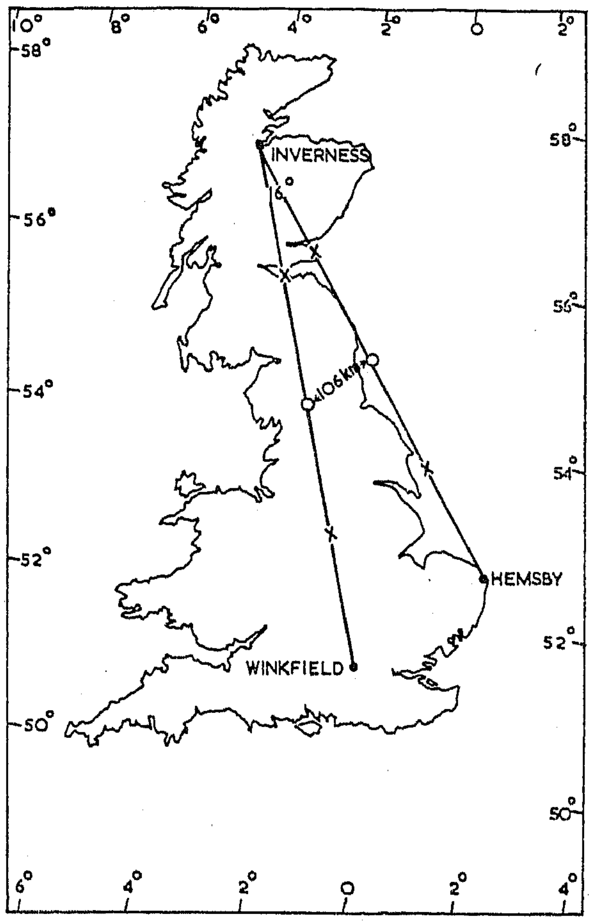


Fig. 1.—Plan of transmission paths.
× Second-order ionospheric reflection.
○ Second-order ground reflection and first-order ionospheric reflection.

intervals of 30sec on the first-order echo, irrespective of the signal amplitude, provided that this was sufficient for recording. Alternate measurements on the second-order echo were also made at 30-sec intervals when possible (i.e. four observations per minute in all), but the relatively low signal strength of this echo often limited the number of observations made on it. No attempt was made to synchronize individual observations at the two direction-finders, but information regarding the pulse patterns received was exchanged by telephone, from time to time, while the measurements were being made.

(3) RESULTS

(3.1) General

At the outset of the experiments it was found that much more diverse results, as between the two separated transmission paths, were in evidence than in previous experiments over closer paths. It was, in fact, sometimes impossible to identify with certainty corresponding echoes in the patterns observed at the two stations, and this was particularly the case when the pattern was complicated by the presence of reflections from the F1 layer. However, as the period of the experiments was centred near midwinter, this complication was not serious, and in most cases a simple pattern was obtained, consisting of 1E, 1F and 2F echoes. At times when the working frequency was close to the F-layer m.u.f., the characteristic high- and low-angle ordinary and extraordinary components were seen, and with simple layer conditions could be readily identified.

(3.2) Comparison of First-Order Results

A comparison of the slow bearing variations of the 1F echo measured at the two receivers over the same periods shows results in marked contrast to those of the previous experiments, with comparatively small separation of the reflection points. In the present case the variations appear entirely dissimilar, and on only a few occasions during the whole of the experiments has it been possible to identify corresponding features in the results. However, the magnitude of the slow changes varied considerably from time to time, and there was a general tendency for the larger variances to occur together at the two stations, although there were some striking exceptions to this behaviour. There was also a general correspondence between the magnitudes of the rapid fluctuations at the two places. In the angles of elevation, large slow changes—presumably due to overall ionospheric height variations—were also often well correlated.

In previous work[2] a quantitative study of the correspondence between the layer tilts at two separated reflection points was made by calculating the correlation coefficient between the directional variations at the two stations (smoothed to eliminate the rapid fluctuations), for each of a number of periods of about an hour. The north–south and east–west components of the tilt were studied and were found to show very similar statistical properties, but it was shown that, in general, there is no correlation between their instantaneous magnitudes. The correlation between corresponding components at two separated reflection points was also found to be the same for north–south as for east–west. In these earlier experiments, reflection-point separations of 13 and 46km were used, and it was estimated from the results that the correlation would fall to zero at a distance of the order of 50km; it was conjectured that it might be negative at greater distances. In the present experiments the separation was 106km, but strict comparison between corresponding components of the tilt was not possible, because the transmission paths were not parallel, but inclined to one another at an angle of 16° (Fig. 1). The actual correlation studied was that existing between the slow changes in the bearings on the first-order echo, which corresponded to tilts at the two reflection points about axes inclined at this angle of 16° to each other. Accurate correlation between the tilts about the orthogonal axes could not be obtained by comparison of the variation in the angles of elevation, because these are also affected by changes in the height of the ionosphere. For example, a large correlation coefficient obtained between the angles of elevation could be produced by large overall-height changes or by true correlation between the ionospheric tilts, but the two effects are inseparable. (In the previous experiments, mentioned above, the distances of transmission were short enough for the effects of height changes to be negligibly small in comparison with the effects of tilts.)

For each hourly period the correlation coefficient between the bearing changes was calculated for zero time displacement of the results, and also for relative time displacements of up to 10min each way. The resulting coefficients provide cross-correlograms similar to those used in previous experiments of this kind.[2] In contrast to previous results, these showed no well-defined general form, and the three examples given in Figs. 2(*a*), 2(*b*) and 2(*c*) illustrate their diverse nature. Table 1 shows the results of an analysis of the correlation coefficients obtained from 30 sets of results for various time displacements. For each time displacement the mean correlation coefficient, ρ and its estimated standard deviation, σ, are given.

There is a tendency for the correlation to be negative for time shifts around zero, but these values are seen to be scarcely significant at the 5% level. Thus, for all practical purposes the tilts at reflection points with this separation (106km, N 49° E) may be regarded as uncorrelated. The rising values of ρ for increasing time shifts either way indicate a tendency [exemplified in Fig. 2(*c*)] towards periodic bearing changes with a period of 20–30min.

Fig. 2(*d*) is a correlogram relating to observations taken during 1½ hours just after ground sunset on a day when a

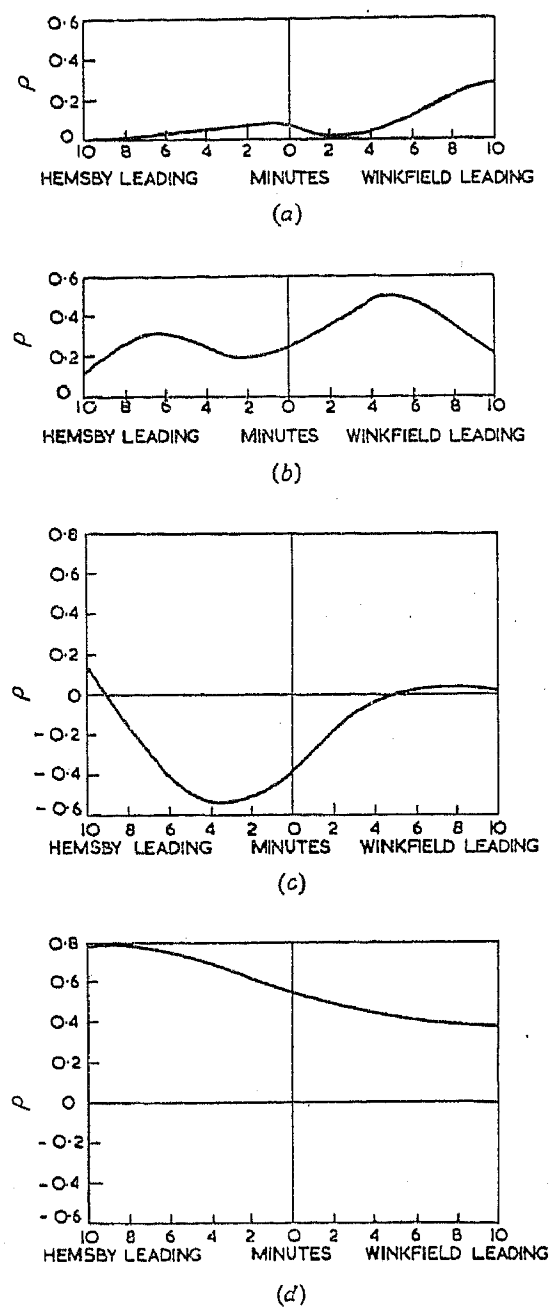


Fig. 2.—Correlograms for 1F bearings at Winkfield and Hemsby.

(a) 1030–1145 U.T., 21st October, 1952; 7·74 Mc/s.
(b) 1335–1500 U.T., 23rd October, 1952; 7·74 Mc/s.
(c) 1550–1655 U.T., 3rd March, 1953; 6·58 Mc/s.
(d) 1635–1810 U.T., 12th November, 1952; 5·58 Mc/s.

Table 1

Correlation Coefficient between 1F Bearings at Winkfield and Hemsby

	Hemsby leading by				Winkfield leading by		
	10 min	6 min	3 min	0	3 min	6 min	10 min
ρ	0·12	−0·05	−0·09	−0·10	−0·02	0·09	0·15
σ	0·07	0·05	0·05	0·05	0·05	0·05	0·06

marked sunset effect was observed. The positive correlation shown in this case (which is not included in Table 1) is accounted for by a systematic ionospheric tilt over the sunset period on this day. The tilt was in the expected sense, i.e. down towards the west, and produced bearing shifts of $2\frac{1}{2}$° towards the west at each direction-finder. Recovery to normal was effected about $1\frac{1}{2}$ hours after ground sunset. The maximum correlation occurred at a time displacement of about 10min (Hemsby leading); the time difference corresponding to the longitude separation of the reflection points for the two trajectories was $4\frac{3}{4}$min, in the same sense.

(3.3) Comparison of Second-Order Results

At the start of the experiments it soon became apparent that there was a marked qualitative difference between the behaviour of the 2F echoes at the two receiving stations, in respect of the amplitude of the 2F echo relative to that of the 1F echo. At Hemsby, the 2F echo, although generally much weaker than the 1F echo, could usually be seen, provided that the working frequency was not too high, and satisfactory measurements could be made on it. At Winkfield, on the other hand, the 2F echo was so weak that measurements of its direction of arrival could be made only infrequently. It was also more spread out in time of arrival than at Hemsby, giving the appearance of a scattered echo. When the echo was strong enough, it could be seen that the direction of arrival was varying rapidly, in a manner characteristic of wave-interference conditions. At Hemsby, the echo was in every respect more like a first-order reflection. This was particularly noticeable when the working frequency was near the m.u.f. for the second-order ray, when the characteristic increase in equivalent path and angle of elevation was well marked. The appearance of an extra component, which was probably the high-angle, or Pedersen, ray and which coalesced with the main echo as the m.u.f. condition was reached, was also sometimes noted. At Winkfield, on the other hand, although well-defined multiple components in the 2F echo were occasionally seen, this m.u.f. effect was never clearly observed.

It was thought possible that the difference between the relative 1F and 2F amplitudes at Winkfield and Hemsby might be due to the polar diagram of the transmitting aerial, which was a vertical rhombic, since the bearings of the two receivers from Inverness differed by 16°. However, calculations showed that the ratio of the field strengths of radiation emitted at the angles of elevation appropriate for 1F and 2F transmission was substantially the same for either bearing. There were also no differences in the receiving aerials which would account for the different behaviour at the two sites, and it therefore appears that this can be attributed to a difference in the reflection properties of the land and sea at the mid-points of the trajectories.

As the experiments progressed, more 2F measurements became available, so that a quantitative comparison of the variations in direction of arrival could be made. It was now found that, although the appearance of the Hemsby results (plotted against time) was generally more coherent than the corresponding ones at Winkfield, the total variances were on the whole not very different. Fig. 3 shows an example of these results, the angles of elevation and the bearings being shown for both the first- and second-order echoes. The slower fluctuations, due to tilts, are much more evident in the Hemsby 2F results than in those for Winkfield.

For the purpose of numerical analysis the results have been divided up into periods of about an hour, and for each period the variance of the individual bearings about the mean for the period has been calculated. For most of the 1F results a fairly accurate division into the "slow" and "rapid" components of variance could be made, and this was done wherever possible. Since the number of 2F observations per period varied considerably, a single overall variance estimate for each type of bearing was made by pooling all the observation periods for which that variance had been calculated. (In this process the individual errors were still referred to the period mean.) The first two lines of Table 2 show the results of this analysis, the variances being given in degrees squared.

The last two lines of the Table show the results of a similar calculation using only selected periods during which results were obtained at both Winkfield and Hemsby on both 1F and 2F echoes. These provide a fairer basis of comparison of the results at the two stations, and, as might be expected, the corresponding

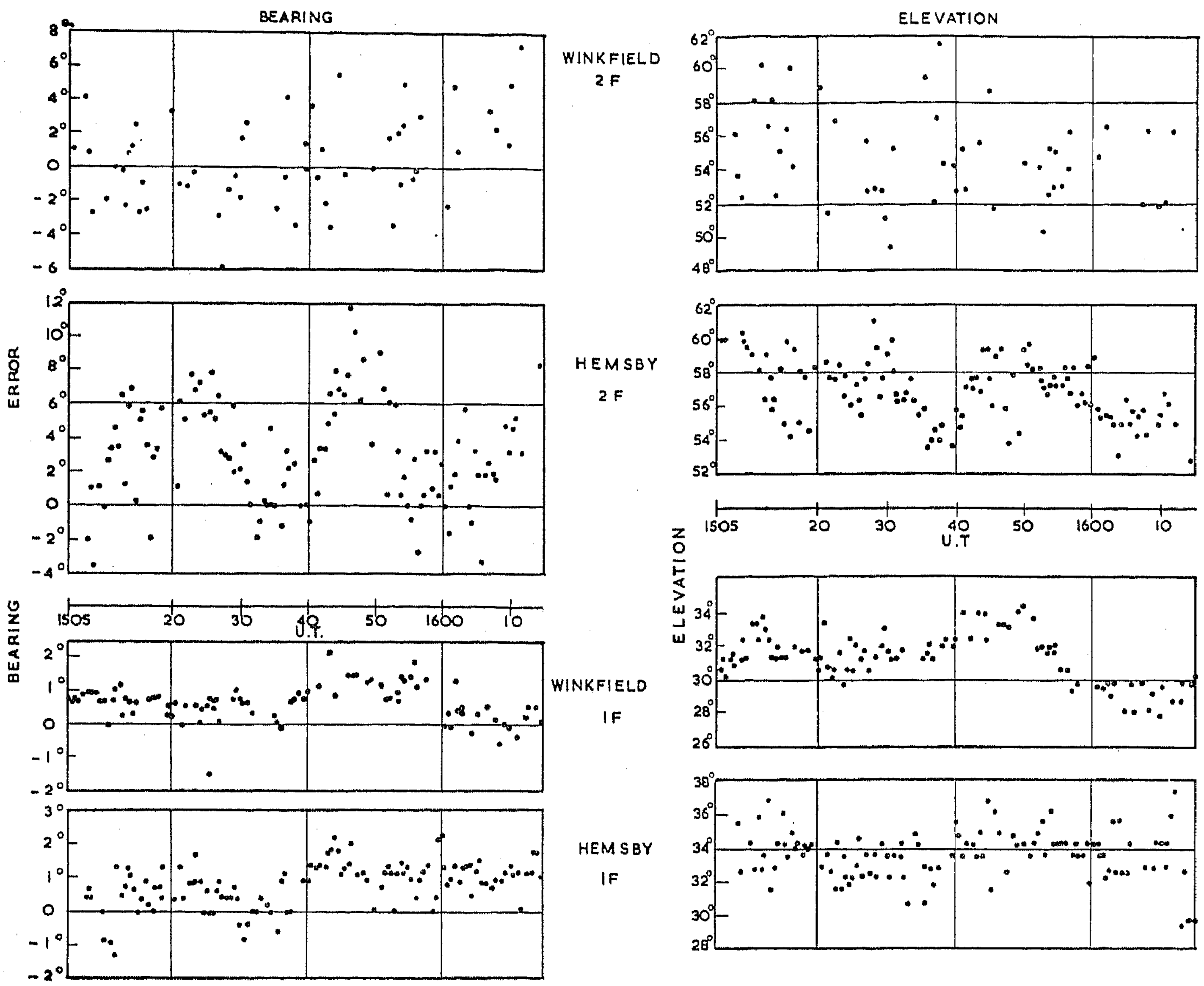


Fig. 3.—Simultaneous measurements of bearing and elevation angle at Winkfield and Hemsby.

First- and second-order F reflections.
20th February, 1953; 5·25 Mc/s.

Table 2

OVERALL VARIANCES OF 1F AND 2F BEARINGS

	Receiver	1F				2F	
		Number of observations	Slow variance	Rapid variance	Total variance	Number of observations	Total variance
			deg²	deg²	deg²		deg²
All results ..	Hemsby Winkfield	7 397 7 344	0·63 0·66	0·26 0·32	0·89 0·98	2 114 1 176	12·0 14·7
Selected periods	Hemsby Winkfield	2 250 2 338	0·73 0·71	0·31 0·35	1·04 1·06	1 318 942	12·2 13·4

variances at Winkfield and Hemsby are in somewhat better agreement than in the previous case. In this case the difference between the 2F variances is only $1 \cdot 2\,\mathrm{deg}^2$, but with the number of observations given this difference is significant at the 5% probability level. Thus it may be concluded that, under the conditions of the experiment, the different behaviour of the land and sea at the mid-points is established, but that it introduces only a small difference in the total variance of the 2F bearings.

In Section 7 a theoretical discussion is given of the relative effects on first- and second-order reflections over the same transmission path, the slow and rapid types of variation being considered separately. Eqn. (14) shows that the rapid component of 2F variance would be expected to be about five times that of the 1F variance if the ground were perfectly smooth, so that the mid-point reflection did not provide any additional variance contribution. Taking the rapid 1F variance as $0 \cdot 3\,\mathrm{deg}^2$ from Table 2, the rapid component of 2F variance would then be only about $1 \cdot 5\,\mathrm{deg}^2$. This is only a small proportion of the total 2F variance, so that the remaining slow component should be clearly seen. In the Hemsby results this is sometimes the case, as in the example shown in Fig. 3, and it can then be said that the mid-point reflection at the sea is almost specular. In general, however, it appears that this reflection contributes some extra variance which tends to obscure the slower lateral deviation

effects, and this process is more marked on the Winkfield results where the mid-point reflection is on the land.

It is not possible to make an accurate division of the average overall 2F variance into rapid and slow components, since in many cases, even at Hemsby, the rapid fluctuations constitute a large enough proportion of the total variance to obscure the slower changes. However, it can be said that the results are not inconsistent with a factor k_S of about 10 [see eqns. (7) and (15)], as expected for uncorrelated tilting at the two ionospheric 2F reflection points.

It may be noted that both at Winkfield and Hemsby the total 2F variance for a particular period shows significant positive correlation with the total 1F variance for the same period. Furthermore, the total 2F variances for corresponding periods at Winkfield and Hemsby are positively correlated, as are the 1F variances (see Section 3.2). Thus, although the experiments show that negligible correlation exists between the instantaneous tilts at points about 100km apart, they also indicate that there is a general correspondence between the magnitudes of the directional changes produced at reflection points over a wide area, of the order of $2 \times 10^4 km^2$ at least.

(3.4) Comparison of Mean Bearing Errors

An analysis has also been made of the mean bearing errors over hourly periods, obtained at the two stations, on the 1F and 2F echoes. At either station a significant positive correlation was found between the 1F and 2F period-mean errors, i.e. a large mean error on the 2F echo was associated with a large mean error of the same sign on the 1F echo. This indicates the existence of real mean tilts over an hourly period, which are, moreover, correlated at the various reflection points along each transmission path. From the range of variation of these tilts, they appear to be not more than about 1° in magnitude. In accordance with eqn. (9) the effect on the bearings is about four times as large on the 2F as on the 1F echoes. Correlation has also been obtained between the period-mean 2F bearings at Winkfield and Hemsby, showing that these mean tilts may exist over the whole area covering the various reflection points shown in Fig. 1.

By plotting the mean bearing errors against time of day, a definite indication of a diurnal trend has been obtained, both on the 1F and 2F echoes. Over the period 1000–1600U.T. there is a swing of the bearings towards the west at a rate of about 0·15deg/h on the 1F bearing and 0·6deg/h on the 2F bearings. This represents a layer tilt changing at the rate of 0·2deg/h, and is in the same sense and of similar magnitude to the effect already reported by Budde.[5] It produces an entirely negligible contribution to the total variance occurring in a period of an hour, and so does not affect the correlations discussed in Section 3.2.

(4) CONCLUSIONS

The following conclusions may be drawn from the experimental results described above:

(*a*) The correlation between closely corresponding components of ionospheric (F-layer) tilt is negligibly small at two points separated by 106km in a direction N 49° E, although this may not hold at times near sunrise or sunset, when large-scale tilting may occur. A daily variation of the tilt in an approximately east–west direction has been found to occur at a rate of about 0·2deg/h.

(*b*) On second-order F-layer transmissions over distances of about 700km, with mid-points on land and on sea, the land behaves as a rougher reflector than the sea for frequencies of 5–8 Mc/s. The effect of this is to increase the rapid directional fluctuations of the echo, but in the case studied this added only a small amount to the total bearing variance, which was caused mainly by ionospheric tilts.

(*c*) The variance of bearings on the 2F reflection at the range of 700km is about 12 times that of bearings on the 1F reflection.

(5) ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of a number of his colleagues, particularly Messrs. A. Ambery and J. Bell, in the experimental work and analysis of the results.

Thanks are due to staff of M.O.12 Department of the Meteorological Office, for valuable assistance in obtaining the site at Hemsby and providing facilities for its use.

The correlograms referred to in Section 3.2 were calculated by the Mathematics Division of the National Physical Laboratory.

The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(6) REFERENCES

(1) BRAMLEY, E. N., and ROSS, W.: "Measurements of the Direction of Arrival of Short Radio Waves reflected at the Ionosphere," *Proceedings of the Royal Society*, A, 1951, **207**, p. 251.

(2) BRAMLEY, E. N.: "Direction-Finding Studies of Large-Scale Ionospheric Irregularities," *ibid.*, 1953, **220**, p. 39.

(3) BRAMLEY, E. N.: "Some Aspects of the Rapid Directional Fluctuations of Short Radio Waves reflected at the Ionosphere" (see page 533).

(4) ROSS, W., BRAMLEY, E. N., and ASHWELL, G. E.: "A Phase-Comparison Method of Measuring the Direction of Arrival of Ionospheric Radio Waves," *Proceedings I.E.E.*, Paper No. 1134 R, July, 1951 (**98**, Part III, p. 294).

(5) BUDDE, W.: "Untersuchungen über die Ausbreitung von Raumwellen," *Zeitschrift für Naturforschung*, 1950, **5a**, p. 683.

(6) ROSS, W.: "Site and Path Errors in Short-Wave Direction Finding," *Journal I.E.E.*, 1947, **94**, Part III, p. 108.

(7) APPENDIX

(7.1) Theoretical Considerations

In considerations of the effect of ionospheric tilts on the direction of arrival of a wave reflected at oblique incidence, it is convenient to resolve the tilt into two orthogonal components with horizontal axes in, and normal to, the great-circle plane of propagation. If the layer at any time is tilted at an angle θ to the horizontal, and the axis of tilt makes an angle ψ with the normal to this plane, the component $\theta \cos \psi$ produces a change in the angle of elevation of the wave and the component $\theta \sin \psi$ produces a change in the bearing.

The relative effects of small layer tilts on first- and second-order reflections will now be examined, taking for simplicity the case of plane-earth transmission; for the distances actually involved the curvature of the earth introduces only minor modifications. If the height of the layer is h and the distance of transmission is D, the component of tilt $\theta \sin \psi$ produces, on the first-order reflection, a change in bearing given by

$$\alpha_1 = (2h/D)\theta \sin \psi \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and the orthogonal component of tilt produces a change in elevation angle Δ_1 given by

$$\Delta_1 = \theta \cos \psi \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now suppose that, for a second-order reflection over the same distance D, the layer tilts at the first and second ionospheric reflection points are specified by (θ', ψ') and (θ'', ψ'') respectively. It may then be shown that the overall effect is a bearing error of

$$\alpha_2 = (2h/D)(\theta' \sin \psi' + 3\theta'' \sin \psi'') \quad . \quad . \quad . \quad (3)$$

and an elevation shift of

$$\Delta_2 = \tfrac{1}{2}(\theta' \cos \psi' + 3\theta'' \cos \psi'') \quad . \quad . \quad . \quad . \quad (4)$$

Thus the tilt at the reflection point nearer to the receiver largely determines the directional changes observed there. Under the same conditions the reflection point at the earth, which is normally half-way between the transmitter and receiver, is displaced along the line joining them by a distance

$$x = (h + D^2/16h)(\theta' \cos \psi' + \theta'' \cos \psi'') \quad . \quad . \quad (5)$$

and normal to this line by a distance

$$y = (h + D^2/16h)(\theta' \sin \psi' + \theta'' \sin \psi'') \quad . \quad . \quad (6)$$

In the present experiments, taking h as 250 km and D as 650 km, it is seen that simultaneous tilts in the same direction of the order of 2° would be required to move the Inverness–Hemsby mid-point, which was nominally on the sea, on to the land.

To compare statistically the magnitude of the angular fluctuations attributable to tilts, on first- and second-order echoes, it may be assumed that θ' and θ'' have the same statistical properties as θ, and that ψ, ψ' and ψ'' vary at random. The resulting variance of α_2 and Δ_2 will depend on the correlation between the tilts at the two reflection points, which, for a 700 km path, are 350 km apart. It is seen from eqns. (1)–(4) that the following relations exist between the variances of α_1 and α_2, and between those for Δ_1 and Δ_2, in the cases specified.

(*a*) No correlation between the tilts:

$$\overline{\alpha_2^2} = 10\overline{\alpha_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$\overline{\Delta_2^2} = 2{\cdot}5\overline{\Delta_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

(*b*) Perfect correlation between θ' and θ'', and between ψ and ψ'':

$$\overline{\alpha_2^2} = 16\overline{\alpha_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\overline{\Delta_2^2} = 4\overline{\Delta_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

(*c*) Perfect negative correlation between $\theta' \sin \psi'$ and $\theta'' \sin \psi''$ and between $\theta' \cos \psi'$ and $\theta'' \cos \psi''$:

$$\overline{\alpha_2^2} = 4\overline{\alpha_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$\overline{\Delta_2^2} = \overline{\Delta_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Thus the ratio of the variances on second- and first-order bearings, due to tilts, can be between 4 and 16, with 10 as the value to be expected for negligible correlation at the reflection points.

It has been pointed out that the directional deviations on a single echo, caused by ionospheric tilts, are always accompanied by more rapid fluctuations, produced by wave-interference effects. Thus the total variance comprises two components, and for a first-order echo at moderate ranges, such as used in these experiments, these can be generally separated by inspection of the results plotted against time.

The major cause of the rapid variance on a first-order ray is believed to be the presence of ionospheric irregularities, which convert an incident plane wave into a diffuse reflection consisting of an angular distribution of energy. In oblique-incidence transmission this angular distribution may be modified by random site errors, i.e. by re-radiation from objects on the ground at distances of up to several kilometres from the receiver,[6] but for short distances of transmission such as are involved here, the angular spread of the downcoming wave itself is probably of major importance in the bearing fluctuations.[3] If now a signal, after a single reflection at the ionosphere, and consisting of a certain angular power distribution determined by the ionospheric irregularities, is reflected at a perfectly smooth earth, this angular distribution will remain unchanged as it impinges on the ionosphere at the second reflection. After the second reflection the distribution will have a wider spread, and if the scattering properties of the ionosphere are the same for each reflection, it may be shown that for narrow distributions the mean-square angular deviation has twice its previous value. This will apply at a given angle of incidence, and since different angles of incidence occur for first- and second-order reflections over a given distance of transmission, this relation will be modified if the angular distribution produced by reflection is a function of the angle of incidence. However, if it is assumed as a first approximation that this distribution is substantially the same for the angles of incidence applicable for first- and second-order reflections (about 55° and 35° respectively, at Winkfield), the angular deviations in the plane of propagation will be related by

$$\overline{\Delta_2^2} = 2\overline{\Delta_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

This relation will not be directly applicable to the measured variations in the angle of elevation, which are known[4] to be affected by the random site errors previously mentioned. The effects are particularly marked at the lower angles of elevation, and thus the 1F measurements will show greater fluctuations, relative to the 2F measurements, than are indicated by eqn. (13).

For bearings, the 2 : 1 increase in variance in going from first-order to second-order reflections will probably not be much affected by site errors, but it will be modified by the fact that the angular distribution producing bearing fluctuations is in a plane inclined at the angle of elevation δ to the horizontal, whereas the fluctuations are measured in the horizontal plane itself. Thus they are magnified by a factor sec δ, and under the above conditions the relation between the rapid components of the bearing variance is

$$\overline{\alpha_2^2} = 2(\cos^2 \delta_1/\cos^2 \delta_2)\overline{\alpha_1^2} \quad . \quad . \quad . \quad . \quad (14)$$

With the average values of δ_1 and δ_2 occurring in the experiments, this gives a variance ratio of about 5 for the rapid bearing fluctuations.

The effect of reflection by a rough surface at the mid-point, instead of by a smooth earth as assumed so far, would be to increase the numerical factors on the right-hand side of eqns. (13) and (14). For instance, if the earth were of the same degree of roughness as the ionosphere, these factors would both be 3 instead of 2.

The total variance, v_2, of a second-order echo may be expressed as the sum of the slow and rapid variances discussed separately above. If these are represented by v_{2S} and v_{2R} respectively, the relation may be written

$$v_2 = v_{2S} + v_{2R} = k_S v_{1S} + k_R v_{1R} \quad . \quad . \quad . \quad (15)$$

while the total variance, v_1, for a first-order echo is given by

$$v_1 = v_{1S} + v_{1R} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

These equations apply to either bearings or angles of elevation; k_S and k_R are numerical factors whose values will be of the order indicated above in eqns. (7)–(12), (13) and (14) respectively.

[The discussion on the above paper will be found overleaf.]

DISCUSSION ON THE ABOVE FOUR PAPERS BEFORE THE RADIO SECTION, 2ND MARCH, 1955

Mr. W. Ross: A short time ago a colleague commented that there had been no substantial advances in high-frequency direction-finding for twenty years. He meant that we are still using the same U-type Adcock direction-finder as was used 20–25 years ago. A closer examination of the situation, however, shows that although there may not have been much change in the actual equipments, we do know a great deal more about direction-finding, and are able to make better use of the direction-finder now than we could then. I only wish that my friend were present this evening so that he might have an opportunity of admitting that, after all, we have not really been wasting our time for twenty years.

Taking Mr. Bramley's second paper first, I consider that it marks another step forward in our understanding of the slowly varying undulations which occur in the ionosphere. Having already shown that ionospheric tilts are correlated over distances of 10–20 km, he has now increased the distance between observing stations to the point where they are uncorrelated. This information is very important in the application of high-frequency direction-finding so that we can understand better how, for example, to place direction-finders in order to achieve the maximum fixing accuracy.

His results on multiple echoes are also of great interest. Once we began to understand that irregularities in the ionospheric layers caused fluctuations in bearing, it became clear that we should have to consider the influence of the ground, which comes into the picture when dealing with multiple reflections. It was immediately obvious when the experiments that Mr. Bramley describes were started several years ago that multiple echoes were very inferior to single reflections in their directional characteristics. Mr. Bramley has been able to show that, on substituting for normal irregular ground a piece of reasonably flat sea as the "bounce" point of the multiple echo, there is an appreciable improvement in its directional characteristics. It might perhaps be commented that the results referred to are not necessarily typical of a true sea reflection case, because the mid-point of the path is fairly close in-shore and undoubtedly the ground on shore must have had some effect. Mr. Bramley's results emphasize, however, that in practical direction-finding, if one can select from the composite received signal a component which has not struck a rough patch of earth, one is likely to achieve better direction-finding accuracy.

In Mr. Bramley's first paper he shows that, while the rapid fluctuations tend to be smaller than the deviations due to lateral deviation when dealing with steeply incident waves, the same is not true for waves of small angles of elevation, so that at the greatest distances it is the rapid fluctuations which predominate.

I would draw attention to Mr. Bramley's reference to site errors, as one of the causes of the "coning" of the ray. Site errors are responsible for some of the quite rapid fluctuations as well as being the cause of more systematic errors in bearing.

Dr. Bain's paper was of great practical interest, since for most applications one is dealing with continuous-wave signals. He has demonstrated quite clearly the length of time needed to average out the rapid fluctuations. It has always been appreciated that a bearing averaged over a period would be better than a single spot reading. Dr. Bain has shown that, so far as the rapid fluctuations are concerned, a period of $\frac{1}{2}$ min or so is the sort of time that is necessary. If we are to remove the still slower fluctuations due to lateral deviation it is necessary to go to a much longer period of averaging; it is no good merely extending the half-minute to a few minutes. This can be very important in deciding, for example, how far one is able to load up a direction-finding network.

Mr. Bowen's paper is really two papers in one, the first dealing with a statistical analysis of bearing errors, and the second with site resistivity effects. The first part is probably more important and more interesting, because one must understand how the various components of the error contribute to the overall bearing error before one can attempt to improve bearings by any system of averaging. It is important in another connection, because when it comes to using direction-finding for position-fixing, it is necessary to be able to assess the reliability of the individual bearings contributing to the fix. This is what is commonly called "bearing classification."

Bearing classification in the past has often been regarded as something of a mystery, a mystery to which comparatively humble operators with no detailed knowledge of propagational processes were supposed to hold the key, being gifted with some sixth sense not possessed by trained engineers. In fact, they have no access to any information that is not present in the observations themselves. To be able to analyse the bearings and decide what are the important contributions to the total bearing error is a step forward towards a logical system of classifying bearings, i.e. of assessing *a priori* their probable accuracy.

The second part of the paper, dealing with resistivity effects, I must admit to finding less valuable. I am quite certain that variations in resistivity over a site are factors which should not be ignored in selecting sites, but I do not think that Mr. Bowen has given a convincing account of precisely how they affect bearings. His analysis is perhaps a little too simple in that it deals with changes taking place over distances of the order of a wavelength or even less by a method which is rather analogous to geometrical optics, instead of from a wave analysis point of view. Mr. Bowen himself indicates that the deductions made from the analysis do not tie up very closely with fact.

The overall impression made by these papers is that there is still one outstanding requirement in high-frequency direction-finding. We now know a good deal about the ionosphere and its effects, and we understand the causes of the various fluctuations in bearing. I think we would all agree that there remains one influence we should like to remove, and that is the effect of the site. If we could have a direction-finder that was more immune to the influence of its surroundings it would be a major step forward.

Mr. J. F. Hatch: Although the accuracy of commercial direction-finders commonly in use is generally of a high order, it has been realized for some time that the errors in the instrument itself, including the aerial system of, say, a U-Adcock on a very good site, represent a small proportion of the total errors of the received wave under conditions of wave interference.

An attempt was made during the war to take bearings on various components of rays arriving from enemy signals, and although some quite elaborate equipment was designed, we were handicapped by lack of knowledge of the accuracy to be expected from the components after they had been separated.

I was particularly interested in the results of time averaging given in Dr. Bain's paper, and I can confirm these results from experiments recently carried out on a commercial direction-finder using an electronic method of counting bearings and averaging them over various periods both on continuous-wave and keyed Morse signals.

On recent measurements a time interval of 0·72 sec was used between adjacent readings. The reduction of variance was found to be fairly constant for stations whose distance from the receiver

varied between a few hundred and several thousand kilometres. In Mr. Bowen's paper, time averaging is mentioned in the process of observing a bearing over a period of 30 sec. I should like to know if this refers to a manually operated direction-finder using a swing bearing technique or a cathode-ray direction-finder using a cursor.

It has been my experience that averaging a bearing by either of these methods is liable to error, because the operator is quite often influenced by the position of the pointer, or the cathode-ray-tube trace in the latter portion of the observational period, and not the mean position over the whole period. This may explain some of the large variances due to operator error.

Several methods of time averaging have been tried in recent years, by continuous recording of the pointer position on a tape, or the more modern method of counting degrees on electronic counters, and dividing the total by the number of bearings taken in a given time. In fact, with this technique it would be possible to check the ability of an operator to estimate an average bearing over any given period, but better still to design an automatic method of bearing presentation, not necessarily to replace the operator but to assist him.

Dr. R. L. Smith-Rose: Reading these papers I was reminded of the first contact that I had with the subject of radio direction-finding through one of the earliest, if not the first of the papers on the subject presented before this Institution in 1920 by H. J. Round.* The author gave an account of his experience in developing direction-finders during the First World War, and I found it interesting to look back at that paper to obtain an appreciation of exactly how the subject has moved over the intervening 35 years. One of the first things that interested me was that there is in the paper a graph of bearing fluctuations occurring due to changes in the ionosphere—although this was not stated, of course, at that time—which might well have done for one of the illustrations in the present papers. The bearings were varying over 2° or 3° in a matter of two minutes, but the conditions were, of course, quite different.

Another point which intrigued me in this earlier and classical paper was that quite clearly the workers of those days had in mind accuracies of 1° for their instrument. Indeed, it is specifically stated as the probable accuracy; and in the discussion of that paper there is an interesting account by Admiral Sir Henry Jackson, who was First Sea Lord at the time, of the manner in which a change in bearing of $1\frac{1}{2}°$ on transmissions from the German Fleet was the deciding factor in his final resolution to send the British Grand Fleet to sea, which resulted in the Battle of Jutland. I wonder whether the authors of these four papers would like to risk their reputations in making such a decision under similar conditions!

Referring to Mr. Bowen's paper, I should like to ask whether the diagram showing the conductivity contours of the ground is typical of the sort of sites that we ourselves have used in extensive experimental work on direction-finders. I cannot think that we found the conductivity of sites varying over a range of more than 2 : 1, and I wonder how the author's observations of conductivity were made in such fine detail as to enable him to set the contours so close together. I also wonder whether it is adequate to take into account a depth of 5 ft in determining the conductivity of the surface.

I should like to ask Mr. Bramley a question on the relation between auto-correlation functions in time and distance. I gather that if observations are taken at intervals of more than 5 sec there is no auto-correlation in time. Now in that space of 5 sec the particular irregularity looked for in the ionosphere will have drifted on a distance of the order of one or two kilometres. It would thus appear possible to predict that the lack of correlation in time of 5 sec corresponds to a lack of correlation in distance of a few kilometres.

* "Direction and Position Finding," *Journal I.E.E.*, 1920, **58**, p. 224.

Finally, I should like to sympathize with Mr. Bowen in having to use ordinary operators in making his observations. We learnt from bitter experience many years ago that it is very risky to attempt to have accurate scientific observations made in this way, unless adequate means are available to check the results. The personal error resulting from the use of such observers is often the limiting factor in attaining accuracy in radio direction-finding; and it is not always easy to separate such errors from those due to the instrument and to the vagaries of the ionosphere.

Mr. D. W. G. Byatt: Dealing first with Mr. Bramley's paper on fluctuations over two transmission paths, the large total variance of the 2F reflections is interesting. The fact that the bearing error due to tilts is shown to be inversely proportional to the distance of the transmitter, and is also largely controlled by the nearness of the last ionospheric reflection point, has been my experience in recent experiments.

Some of the effects of wide fluctuations of relatively near transmitters compared with distant ones can be accounted for by the decrease in polarization protection of a U-Adcock for high-angle rays, but I believed, as has been shown, that this was not the main effect. In the set of observations shown in the paper there seems to be some correlation between bearing fluctuations and change of elevation of the ray. This may, however, be an isolated case and insignificant.

Referring to Dr. Bain's paper, I had available a continuous set of recordings of bearing taken at the rate of $12\frac{1}{2}$ per second on similar equipment during 1945. This also indicated strong auto-correlation for an interval less than about half a second.

In dealing with possible improvements in operational high-frequency direction-finding, the two main sources of error to be tackled seem to be those due to wave interference and observation. Automatic indication of bearing, although difficult to make comprehensive for the numerous types of transmission in the high-frequency bands, should go a long way to solve the operational problem. Combining this with some form of time-averaging will help the removal of much of the error due to wave interference.

Mr. Bowen's paper is interesting particularly with respect to the effect of ground resistivity on the performance of a grounded U-Adcock. Is it the author's opinion that the effects would be less on a balanced aerial system, say of the loop type?

I should like to remark on the fact that all the measurements on directional fluctuations have, naturally, been made on either the vertical electric or horizontal magnetic component of the wave. The polarization of the wave from a distant transmitter has random, including circular, polarization on arriving at the direction-finder, and it is conceivable that when wave-interference effects in one plane of polarization cause large bearing errors, as well as fluctuations, the waves in the other plane may be strong and normal.

Measurements have been made that show considerable reduction in fading if both the horizontal and vertical components of the wave are received, the improvement being comparable to that achieved with spaced diversity. Is Dr. Bain of the opinion that useful reduction in bearing fluctuations could also be made if both components of the incident waves could be sampled?

Mr. F. J. M. Laver: My interest is not in high-frequency direction-finding as such but in aerials for high-frequency point-to-point radio systems. In such systems there is much to be said for spending money on the receiving aerial rather than on increasing transmitter power. Thus, increasing the transmitter power increases mutual interference, but improved receiving aerials help considerably to decrease it.

An important question is, How far is it practicable to narrow the beam of the aerial? There seems to be some danger that, if the beam is too narrow, fluctuations in direction of arrival may take the signals outside the beam and result in loss of signal strength. One can imagine a pencil-beam array automatically following variations in the direction of signal arrival, but I conclude that the rapid fluctuations which Mr. Bramley describes are the response of his direction-finder to the diffusion of the signal energy rather than actual variations of some resultant ray. If that is so, it is difficult to visualize what fluctuations in direction the array as a whole should be steered to follow.

Moreover, I should like the authors' opinion on whether the phase incoherence which ionospheric irregularities introduce prevents the large arrays from having high directivity. Thus, bearing correlation is lost at a spacing of 400 m, which is comparable with the dimensions of a large linear array which might be used to produce a pencil beam for high-frequency reception. Presumably, the phases of signal currents induced in the individual aerials of such an array are more or less uncorrelated, and will not combine in the intended way to produce a pencil beam. Does this mean that 400 m represents the practical limit of size for a high-frequency receiving aerial, bearing in mind that such an aerial would be most useful when conditions were disturbed?

The spacing of 400 m is measured transversely; has the author a similar figure measured along the direction of arrival?

Mr. Bowen mentions rare occasions of low signal strength and wide deviations. Can he say how often wide deviations occur and how large they are? We have heard on other occasions of very large deviations in the apparent direction of arrival of signals. Have the authors observed any large sustained deviations from true bearing?

Dr. Bain's Fig. 1 gives two curves of bearings of signals from Leipzig which differ considerably; is any reason known to account for this difference? The lower curve interests me particularly, because it seems to have a number of variations of similar character; these consist of a positive spike immediately followed by an equal negative spike, and I should like to ask whether they are characteristic of wave-interference effects.

Mr. D. S. Palmer: The variance ratio test used by Mr. Bramley in discussing the Hemsby and Winkfield results is valid only if successive members of the populations tested are independent. Here 2F bearing errors at Hemsby and Winkfield are compared and their numbers are taken as 1 318 and 942. The variance ratio of 1·2 comes out significant; but as the Hemsby figures at least are highly correlated, the number of effective independent observations is much less. The period containing one independent observation on the variance is approximately $2\int_0^\infty \rho^2(t)dt$ where ρ is the auto-correlation.

We can estimate the number of independent points in the Hemsby figures illustrated in Fig. 3 of Mr. Bramley's first paper. The slow variation passes through the mean about four times an hour, and if we assume an autocorrelogram $e^{-t^2/2t_0^2}$, t_0 by Rice's equation is $1/4\pi$. Take the long-term variance as 80% of the whole so that the autocorrelation is $0\cdot 8e^{-t^2/2t_0^2}$, together with an extra element for small t, quickly falling away; the integral gives $5\frac{1}{2}$ independent points an hour in place of 100 in the diagram, and on the same proportion would give 72 independent points in all instead of 1 318.

If we use the approximate result that the variance ratio for large numbers n_1 and n_2 is normally distributed with mean 1 and standard deviation $\sqrt{\frac{2(n_1 + n_2)}{n_1 n_2}}$, we get standard deviations of 0·06 for 1 318 and 942, but of 0·17 for 72 and 942, assuming that there is no coherence at all in the Winkfield results. The ratio of 1·2 is thus significant in the first case but not significant when allowance has been made for the internal correlation of the Hemsby results.

Dr. G. J. Phillips: I believe that the Appendix to Mr. Bramley's first paper will prove a very valuable contribution to our understanding of the mechanism behind fading and its associated fluctuation effects. A point of interest in this paper is the way in which the directional fluctuations increase as the radio frequency is lowered. Thus both theory and experiment indicate, for near-vertical incidence, a degree of fluctuation varying inversely as the square of the frequency in the h.f. band. At which point does the law cease to hold at lower frequencies? Experiments at medium frequencies show fairly large fluctuations (of the order of 8° at 1 Mc/s), but these are not quite as large as would be expected from the law. This is presumably because the reflection point at medium frequencies would generally lie within the main disturbing region, whereas the law is expected to hold well only if the disturbing region is entirely below the point of reflection.

Concerning instrumentation, has a continuous recording of phase difference between two aerials been considered for investigating variations in direction of arrival?

Dr. H. G. Hopkins: My remarks are mainly concerned with Mr. Bowen's paper, in which I would welcome elaboration of some points of detail. It is not stated whether the data refer to observations made visually or aurally; if aural-null observations are included, do the results show any spurious correlation between successive bearings due to the operator, particularly when the swing about the minimum is large? Such an effect was reported by Ross (*Journal I.E.E.*, 1947, **94**, Part IIIA, p. 722), and it would be of interest to know if the present analyses provide qualitative or quantitative confirmation.

It would be helpful if Mr. Bowen could indicate how the variance due to residual octantal errors (Section 2.4) was extracted from the data. The discussion in Sections 3.2 and 3.3 shows that the composite wave-interference and observer variance was derived by allocating 2 deg^2 for distant site and instrumental errors, although the analysis provides no independent confirmation of this figure. It therefore appears that the logical inference should be that wave-interference and observer variance is less than 5·8 deg^2 rather than "about 5 deg^2," as stated in Section 3.3. The practical distinction between these two statements is, of course, small.

The analysis of the variance of bearings is of importance in arriving at objective methods for the classification of bearing accuracy. The bases for such methods, at any rate in the h.f. band, have been known for a number of years, yet subjective classifications relying on the judgment of the operator still exist. It is to be hoped that the objective classification systems will find more widespread use in the future.

Dr. J. A. Saxton: In his first paper, Mr. Bramley suggests an explanation of the rapid directional fluctuations of waves reflected at the F-region, based on the irregularities known to exist in the density of ionization in the E-region. From his results Mr. Bramley deduces that the "scale" of these irregularities is of the order of 500 m. The corresponding value used by Bailey and his colleagues in interpreting their observations on long-distance v.h.f. transmission by means of ionospheric scattering is 100 m: similar work carried out recently in this country, by the General Post Office and the Department of Scientific and Industrial Research in collaboration, appears to indicate that, for a considerable fraction of the time at least, the scale of the density fluctuations responsible for forward scattering at v.h.f. must be very much less than 500 m, and probably not much more than

10 m. Has Mr. Bramley made sufficient measurements to be able to say what limits are likely to encompass the range of variations in the scale of E-region ionization irregularities?

Mr. S. D. Heesom: Mr. Bowen's paper, dealing with the earth resistivity tests, shows a most impressive diagram, and I wonder whether he can tell us what effect a change in frequency would have on the shape of the contours.

The papers report results of useful investigation, but, as an engineer interested in point-to-point communication, I wonder if any of the results obtained indicate the best layout of aerials for space diversity reception. To reduce the effect of fading due to polarization of the sky wave, the outputs of two or three separate aerials are combined; but have the authors any evidence to suggest that an end-on siting of, say, two aerials might give results differing from a broadside siting with the same aerial separation?

A number of speakers have referred to site errors, and I would like to mention, in passing, that I did hear of one high-frequency direction-finding station where a site error was considerably reduced after an old rusty bicycle frame had been dug up.

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Mr. K. C. Bowen (*in reply*): In reply to Dr. Hopkins and Mr. Hatch, the statistical analysis was based on results from manually-operated and visual-indication direction-finders in roughly equal proportions. There is no indication of any marked difference in performance. I agree with Mr. Hatch that, unless the observational technique is correct, only the last five seconds, say, of an observing period may be remembered. Operators are instructed to observe for five seconds and record their bearing data before making further observation: a spurious correlation between successive observations can, as Dr. Hopkins suggests, sometimes be detected. This correlation tends to increase as the swing increases and can create marked anomalies in subjective bearing classification systems based solely on the repeatability of the indicated bearing.

I agree with Dr. Hopkins's comments on the estimation of variance due to wave-interference and operators. The residual octantal error variance was obtained by averaging sets of bearings taken on the same source but on different radio frequencies. The variance attributed to such an average bearing is taken to be the residual octantal error variance, plus the total remaining variance divided by the number of observations in the set.

Mr. Laver is presumably interested in large deviations other than those produced by wave interference or sporadic ionization.* Only one instance of such errors is available from operational data: apparently consistent errors of mean values up to $-20°$ were given by various direction-finders over a considerable period.

Dr. Smith-Rose properly questions the ability to observe the resistivity contour lines so close together: the diagram is interpolated from a series of observations made every 15°, at distances 60, 120, 180 and 240 ft from the centre of the system, using the 4-electrode direct-current technique of measurement.† Measurements made one year later showed very similar contours and confirmed the range of variation of resistivity. Depths greater than 5 ft should be considered when the resistivity is high, and the effects of the vertical gradients referred to in Section 4.3 need investigation.

The site considered is, in range of resistivity, typical of those used in naval organization. The best site has a range of from 2 to 3 kilohm-centimetres, while the worst varies over 10 to 80 kilohm-centimetres. Since there is a greater freedom of choice, geographically, of experimental sites, these may well be as good as the best quoted above.

I agree with Mr. Byatt's implied statement that a balanced aerial system would be advantageous on poor sites: an elevated-H system has been under consideration. It is important, however, to consider very carefully what constitutes a poor site, and, as stressed by Mr. Ross, the analysis I have put forward, with considerable reservation, is not sufficiently precise for such an investigation.

* E.g. Feldman, C. B.: "Deviations of Short Radio Waves from the London–New York Great-Circle Path," *Proceedings of the Institute of Radio Engineers*, 1939, 27, p. 635.

† Morgan, P. D., and Taylor, H. G.: "Measurement of the Resistance of Earth Electrodes," *World Power*, February, 1934, p. 76.

Mr. E. N. Bramley (*in reply*): With regard to the small-scale ionospheric irregularities mentioned by a number of speakers, it should be emphasized that the size quoted is only a parameter in an assumed statistical specification of their magnitudes. In comparing the results with those of other workers it must be borne in mind that the "size" can be defined in various ways, which may account for differences in the quoted values. The existence of very small irregularities such as mentioned by Dr. Saxton is not precluded by the results of the high-frequency measurements; these frequencies may well be little affected by comparison with the v.h.f. signals, for which the size is comparable with the wavelength.

There must of course be a limit to the increase in the fluctuations with decreasing frequency below 3 Mc/s. One reason for this has been mentioned by Dr. Phillips; another contributory factor is the collisional frequency when this becomes comparable with the wave frequency.

It does not appear likely that lack of bearing correlation at points separated by a few hundred metres, on a single mode of propagation, would prevent the full gain of a large highly-directive aerial being realized. This is because in the presence of a strong steady "signal," which has been shown to exist, particularly at the higher frequencies, the phase fluctuations represent only minor irregularities in a wavefront which remains substantially plane over much greater distances.

Dr. Smith-Rose has mentioned the connection between the space and time correlation of the rapid directional fluctuations. A correlation coefficient falling to zero at a time interval of a few seconds, or a distance of a few hundred metres, is consistent with a drift speed of the order of 100 m/sec. The same results could be produced, however, by random movements of the irregularities, with no systematic drifts, and the present results do not enable us to distinguish between the two possibilities.

With regard to the slower variations in direction of arrival, attributed to ionospheric tilts, correlation between the changes in bearing and elevation is not a general feature of the results. When this does occur, it indicates a steady horizontal progression in the reflecting region, and some results of this kind have been described in another paper.*

The significance of the variance ratio obtained for the 2F bearing errors at Hemsby and Winkfield is reduced by the autocorrelation of the Hemsby results, as pointed out by Mr. Palmer, although not to such a large extent as he suggests. For the Hemsby results as a whole the lateral deviation component of variance may be put at about 60% of the total, and the autocorrelation function $(1 - 2t^2/t_0^2)\exp(-t^2/t_0^2)$, with $t_0 = 7\frac{1}{2}$ min, has been shown* to give a good representation of these variations. The integral quoted by Mr. Palmer then gives $2\frac{1}{2}$ min as the period containing one independent observation. The 1 318

* Bramley, E. N.: "Direction-finding Studies of Large-Scale Ionospheric Irregularities," *Proceedings of the Royal Society*, A, 1953, 220, p. 39.

observations were spread over 22 hours, so the effective total number of independent observations is 528. With large-sample theory the standard deviation of the variance ratio is 0·077, and the probability of the observed ratio of 1·1 occurring by chance is then about 8%.

In reply to Dr. Phillips, a system of continuous phase recording would certainly be desirable, but presents considerable difficulties, and a sufficiently accurate equipment, capable of operating on short pulse signals, and particularly in the presence of interfering continuous-wave signals, has not been developed.

Dr. W. C. Bain (*in reply*): I am interested to hear of the results on the time-scale of the rapid fluctuations reported by Messrs. Hatch and Byatt. In reply to Mr. Byatt's point regarding the usefulness of sampling both the vertically and horizontally polarized components of the incident radiation, I believe that it would be valuable to take such bearing measurements on any transmission on which fading is reduced by the reception of both components; this, of course, assumes the existence of a direction-finder operating on horizontally polarized waves with a performance at least equal to that of the standard h.f. Adcock.

In reply to Mr. Laver, changes with time in the type of bearing fluctuation, such as that shown in the Leipzig records on Fig. 1, are very common; these examples have not been associated with particular ionospheric changes but might be ascribed to differences in the relative strengths of ionospheric modes being received. The equal positive and negative spikes in bearing following each other in one Leipzig record are not believed to be especially characteristic of wave interference conditions.

We have never observed any large sustained deviations from true bearing at a time when the critical frequency of the ionospheric reflecting layer was sufficiently high to permit normal propagation between transmitter and receiver, provided that deviations observed at sunrise and sunset on nearby transmitters are excluded.

Messrs. Laver and Heesom have both raised the question of the signal fluctuations at points separated transversely to, and along, the direction of arrival of the signal. The relative magnitude of the effects in the two cases depends on the angular distribution of the signal. When this consists of a single mode of propagation, with a fairly narrow angular spread of directions, the fluctuations are greatest in the transverse direction and least along the line of propagation, i.e. receiving aerials must be separated by a greater distance in the in-line direction to give the same degree of diversity. In multi-mode propagation the effects observed at equally-spaced pairs of points will depend on the relative spreads of the incident radiation in bearing and elevation. Very roughly, if the ratio of the spread in elevation to that in bearing is greater than cot δ, where δ is the mean elevation in the range, then the fluctuations will be greater in the in-line than in the transverse pair, and vice versa. The spread in bearing will usually be quite small, so if the angle of elevation is not too low the fluctuations in the in-line pair will often be the greater in practice, as has been confirmed on at least one transmitter (Arganda, 5·99 Mc/s).

DIGESTS OF INSTITUTION MONOGRAPHS

SOME FUNDAMENTAL PROPERTIES OF NETWORKS WITHOUT MUTUAL INDUCTANCE

621.372.5 Monograph No. 118 R

A. TALBOT, M.A., Ph.D.

(DIGEST *of a paper published in January,* 1955, *as an* INSTITUTION MONOGRAPH *and to be republished in Part C of the* PROCEEDINGS.)

It is easily seen that for positive real values of the complex frequency variable λ, general *LRC* networks without mutual inductance behave exactly like resistance networks. The main object of the paper is to obtain, in an elementary way, a number of results, probably well known but not previously explicitly enunciated, concerning voltage gains and current gains attainable in resistance networks, and using the above principle, to extend these results to *LRC* networks without mutual inductance.

The basic results for resistance networks are contained in the following theorems, in which "voltage gain" means the ratio of output voltage to input voltage, and "current gain" means the ratio of any branch current to the input current.

Theorem 1.—If T is the voltage gain of a 3-terminal resistance network, $0 \leqslant T \leqslant 1$. The extreme values $T = 0$ and $T = 1$ can only occur if at least one branch resistance is zero or infinite, or if the output node belongs to a sub-network hinged at one of the input nodes and not containing the other.

Theorem 2.—If T is the voltage gain of a 4-terminal resistance network, $-1 \leqslant T \leqslant 1$. The extreme values $T = \pm 1$ can only occur if at least one branch resistance is zero or infinite, or if the two output nodes belong to separate sub-networks hinged to their respective complements at the two input nodes.

Theorem 3.—If τ is the current gain in any directed branch of a resistance network, $-1 \leqslant \tau \leqslant 1$; and in the 3-terminal case when the branch begins at the positive input node or ends at the negative input node, $0 \leqslant \tau \leqslant 1$. The extreme values ± 1, or 0, 1, respectively, can only occur if at least one branch resistance is zero or infinite, or if the branch considered is the sole link between two sub-networks to which the input nodes respectively belong; or, in the 3-terminal case, between one input node and a sub-network containing the other.

The special configurations mentioned in the theorems are illustrated in Figs. 1, 2, and 3. The theorems are proved by a

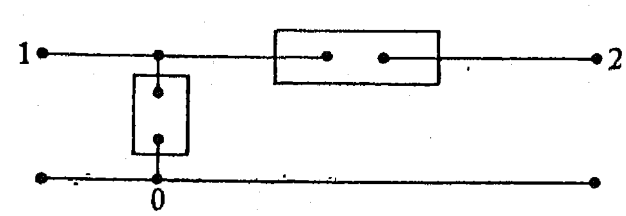


Fig. 1.—Hinged networks: special configuration in Theorem 1.

simple topological examination of the networks, taking into account Ohm's law and Kirchhoff's current law, and they hold even for non-linear networks. By the application of these theorems to an algebraical consideration of the Kirchhoff equations for a network, and of their solution (and without

Dr. Talbot is in the Mathematics Department of Imperial College, University of London.

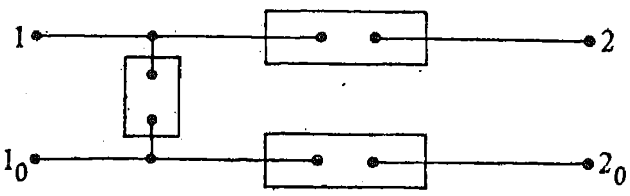


Fig. 2.—Hinged networks: special configuration in Theorem 2.

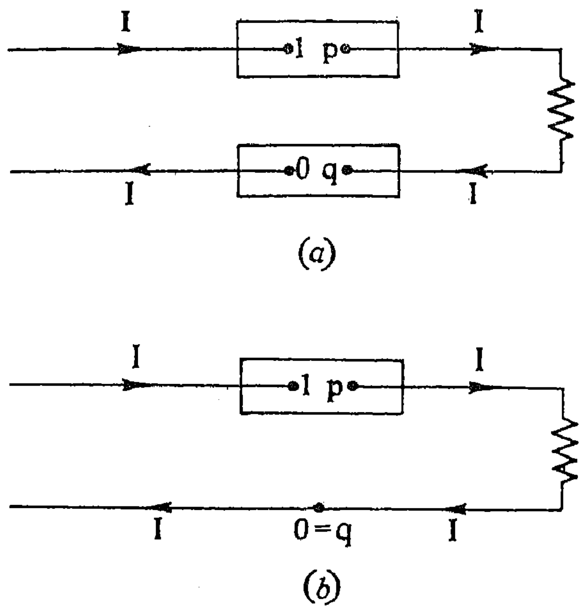


Fig. 3.—Special configurations in Theorem 3.
(a) 4-terminal case, (b) 3-terminal case.

any manipulation of determinants), the following results are deduced:

In a 4- or 3-terminal network, the voltage gain has the form

$$T = \frac{D' - D''}{D}, \text{ or } \frac{D'}{D} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where D is a homogeneous polynomial in the branch resistances, linear in each individual resistance, and with positive coefficients, and D' and D'' comprise some of the terms of D, with the same coefficients.

Similarly, the current gain has the form

$$\tau = \frac{N' - N''}{N_1}, \text{ or } \frac{N'}{N_1} \text{ (in the 3-terminal case)} \qquad (2)$$

where similar remarks apply.

After a recapitulation of well-known properties of general 2- or 4-terminal networks and of reactance networks, the gain theorems together with the above-mentioned principle are used to deduce that if $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$ is the chain-matrix (i.e. matrix relating input quantities to output quantities) of a 3- or 4-terminal LRC network without mutual inductance, then, for $\lambda \geqslant 0$:

$$\text{In a 3-terminal network, } 1 \leqslant \alpha, \delta \leqslant \infty; \; \beta, \gamma \geqslant 0$$

$$\text{In a 4-terminal network, } 1 \leqslant |\alpha|, |\delta|; \; \frac{\beta}{\alpha}, \frac{\gamma}{\delta} \geqslant 0$$

and in fact α, β, γ, δ all have the same sign.

In terms of the admittance and impedance functions, y_{ij} and z_{ij}, we have in the two cases:

$$0 \leqslant -y_{12} \leqslant y_{11}, y_{22}; \; 0 \leqslant z_{12} \leqslant z_{11}, z_{22};$$

$$-y_{11}, -y_{22} \leqslant y_{12} \leqslant y_{11}, y_{22}; \; -z_{11}, -z_{22} \leqslant z_{12} \leqslant z_{11}, z_{22}$$

In all these relations, equality can only occur for $\lambda = 0$ or ∞, unless the network has one of the special configurations mentioned in Theorems 1, 2 or 3.

Moreover, it follows from eqns. (1) and (2) that, for an LRC network without mutual inductance, T and τ have the form

$$\frac{a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n}{c_0\lambda^n + c_1\lambda^{n-1} + \ldots + c_n}$$

where $0 \leqslant a_i \leqslant c_i$ $(i = 0, 1, \ldots n)$ in the 3-terminal case, and $-c_i \leqslant a_i \leqslant c_i$ $(i = 0, 1, \ldots n)$ in the 4-terminal case.

All these results hold, in particular, for RC networks. The additional and well-known fact that the poles of T and τ for RC networks are distinct negative numbers is deduced in the paper from known and more easily established properties of LC networks.

Finally, attention is drawn, in an Appendix, to a paper[10] by Kirchhoff containing a rule, which should be better known, and by which the numerators and denominators of the branch currents in a resistance network can be written down, term by term, by a topological inspection of the network. From this rule explicit expressions are deduced in the paper for the chain-matrix elements of any resistance network, or LRC network without mutual inductance. Expressions for y_{ij} and z_{ij} may be written down from these.

REFERENCES

(1) CAUER, W.: "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," *Archiv für Elektrotechnik*, 1927, **17**, p. 355.
(2) GUILLEMIN, E. A.: "Synthesis of *RC*-networks," *Journal of Mathematics and Physics*, 1949, **28**, p. 22.
(3) FIALKOW, A., and GERST, I.: "The Transfer Function of General Two-Terminal-Pair *RC* Networks," *Quarterly of Applied Mathematics*, 1952, **10**, p. 113.
(4) GUILLEMIN, E. A.: "Communication Networks, II" (Wiley, New York, 1935).
(5) ROUTH, E. J.: "Advanced Rigid Dynamics" (Macmillan, 1892. Fifth edition).
(6) FOSTER, R. M.: "A Reactance Theorem," *Bell System Technical Journal*, 1924, **3**, p. 259.
(7) INGRAM, W. H., and CRAMLET, C. M.: "On the Foundations of Electrical Network Theory," *Journal of Mathematics and Physics*, 1944, **23**, p. 134.
(8) SYNGE, J. L.: "The Fundamental Theorem of Electric Networks," *Quarterly of Applied Mathematics*, 1951, **9**, p. 113.
(9) PERCIVAL, W. S.: "The Solution of Passive Electrical Networks by means of Mathematical Trees," *Proceedings I.E.E.*, Paper No. 1492 R, May, 1953 (**100**, Part III, p. 143).
(10) KIRCHHOFF, G.: "Über die Auflösung der Gleichungen auf welche man bei der Untersuchung der linearen verteilung galvanische Ströme gefurt wird," *Annalen der Physik und Chemie*, 1847, **72**, p. 497.
(11) TALBOT, A.: "A New Method of Synthesis of Reactance Networks," *Proceedings I.E.E.*, Monograph No. 77 R, October, 1953 (**101**, Part IV, p. 73).
(12) BRUNE, O.: "Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency," *Journal of Mathematics and Physics*, 1931, **10**, p. 191.
(13) FEUSSNER, W.: "Über Stromverzweigung in netzförmigen Leitern," *Annalen der Physik*, 1902, **9**, p. 1304; also *ibid.*, 1904, **15**, p. 385.
(14) CAUER, W.: "Über Functionen mit positiven Realteil," *Mathematische Annalen*, 1932, **106**, p. 369.

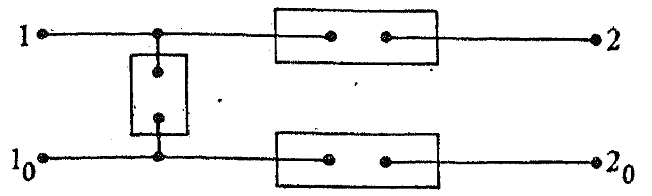
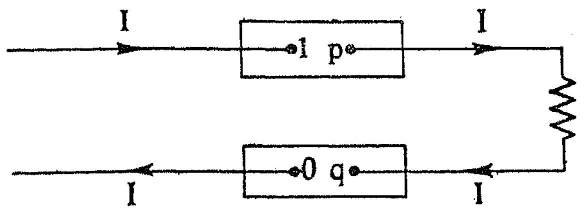
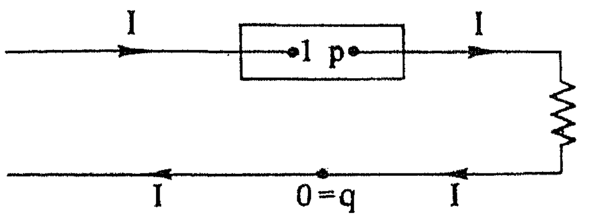


Fig. 2.—Hinged networks: special configuration in Theorem 2.



(a)



(b)

Fig. 3.—Special configurations in Theorem 3.
(a) 4-terminal case, (b) 3-terminal case.

any manipulation of determinants), the following results are deduced:

In a 4- or 3-terminal network, the voltage gain has the form

$$T = \frac{D' - D''}{D}, \text{ or } \frac{D'}{D} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where D is a homogeneous polynomial in the branch resistances, linear in each individual resistance, and with positive coefficients, and D' and D'' comprise some of the terms of D, with the same coefficients.

Similarly, the current gain has the form

$$\tau = \frac{N' - N''}{N_1}, \text{ or } \frac{N'}{N_1} \text{ (in the 3-terminal case)} \qquad (2)$$

where similar remarks apply.

After a recapitulation of well-known properties of general 2- or 4-terminal networks and of reactance networks, the gain theorems together with the above-mentioned principle are used to deduce that if $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$ is the chain-matrix (i.e. matrix relating input quantities to output quantities) of a 3- or 4-terminal LRC network without mutual inductance, then, for $\lambda \geqslant 0$:

In a 3-terminal network, $1 \leqslant \alpha, \delta \leqslant \infty;\ \beta, \gamma \geqslant 0$

In a 4-terminal network, $1 \leqslant |\alpha|, |\delta|;\ \dfrac{\beta}{\alpha}, \dfrac{\gamma}{\delta} \geqslant 0$

and in fact α, β, γ, δ all have the same sign.

In terms of the admittance and impedance functions, y_{ij} and z_{ij}, we have in the two cases:

$$0 \leqslant -y_{12} \leqslant y_{11}, y_{22};\ 0 \leqslant z_{12} \leqslant z_{11}, z_{22};$$

$$-y_{11}, -y_{22} \leqslant y_{12} \leqslant y_{11}, y_{22};\ -z_{11}, -z_{22} \leqslant z_{12} \leqslant z_{11}, z_{22}$$

In all these relations, equality can only occur for $\lambda = 0$ or ∞, unless the network has one of the special configurations mentioned in Theorems 1, 2 or 3.

Moreover, it follows from eqns. (1) and (2) that, for an LRC network without mutual inductance, T and τ have the form

$$\frac{a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n}{c_0\lambda^n + c_1\lambda^{n-1} + \ldots + c_n}$$

where $0 \leqslant a_i \leqslant c_i$ ($i = 0, 1, \ldots n$) in the 3-terminal case, and $-c_i \leqslant a_i \leqslant c_i$ ($i = 0, 1, \ldots n$) in the 4-terminal case.

All these results hold, in particular, for RC networks. The additional and well-known fact that the poles of T and τ for RC networks are distinct negative numbers is deduced in the paper from known and more easily established properties of LC networks.

Finally, attention is drawn, in an Appendix, to a paper[10] by Kirchhoff containing a rule, which should be better known, and by which the numerators and denominators of the branch currents in a resistance network can be written down, term by term, by a topological inspection of the network. From this rule explicit expressions are deduced in the paper for the chain-matrix elements of any resistance network, or LRC network without mutual inductance. Expressions for y_{ij} and z_{ij} may be written down from these.

References

(1) CAUER, W.: "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," *Archiv für Elektrotechnik*, 1927, **17,** p. 355.
(2) GUILLEMIN, E. A.: "Synthesis of *RC*-networks," *Journal of Mathematics and Physics*, 1949, **28,** p. 22.
(3) FIALKOW, A., and GERST, I.: "The Transfer Function of General Two-Terminal-Pair *RC* Networks," *Quarterly of Applied Mathematics*, 1952, **10,** p. 113.
(4) GUILLEMIN, E. A.: "Communication Networks, II" (Wiley, New York, 1935).
(5) ROUTH, E. J.: "Advanced Rigid Dynamics" (Macmillan, 1892. Fifth edition).
(6) FOSTER, R. M.: "A Reactance Theorem," *Bell System Technical Journal*, 1924, **3,** p. 259.
(7) INGRAM, W. H., and CRAMLET, C. M.: "On the Foundations of Electrical Network Theory," *Journal of Mathematics and Physics*, 1944, **23,** p. 134.
(8) SYNGE, J. L.: "The Fundamental Theorem of Electric Networks," *Quarterly of Applied Mathematics*, 1951, **9,** p. 113.
(9) PERCIVAL, W. S.: "The Solution of Passive Electrical Networks by means of Mathematical Trees," *Proceedings I.E.E.*, Paper No. 1492 R, May, 1953 (**100,** Part III, p. 143).
(10) KIRCHHOFF, G.: "Über die Auflösung der Gleichungen auf welche man bei der Untersuchung der linearen verteilung galvanische Ströme gefurt wird," *Annalen der Physik und Chemie*, 1847, **72,** p. 497.
(11) TALBOT, A.: "A New Method of Synthesis of Reactance Networks," *Proceedings I.E.E.*, Monograph No. 77 R, October, 1953 (**101,** Part IV, p. 73).
(12) BRUNE, O.: "Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency," *Journal of Mathematics and Physics*, 1931, **10,** p. 191.
(13) FEUSSNER, W.: "Über Stromverzweigung in netzförmigen Leitern," *Annalen der Physik*, 1902, **9,** p. 1304; also *ibid.*, 1904, **15,** p. 385.
(14) CAUER, W.: "Über Functionen mit positiven Realteil," *Mathematische Annalen*, 1932, **106,** p. 369.

LEAKAGE FLUX AND SURFACE POLARITY IN IRON RING STAMPINGS

621.318.1.042.2 : 621.3.013.5 Monograph No. 116

P. HAMMOND, M.A., Associate Member

(DIGEST *of a paper published in January*, 1955, *as an* INSTITUTION MONOGRAPH *and republished in March*, 1955, *in Part C of the* PROCEEDINGS.)

It is a very surprising fact that the flux distribution around an annular iron ring is substantially constant regardless of the disposition of the magnetizing winding. The problem raised by this constancy of flux is often dismissed by the statement that the iron conducts flux because of its high permeability. This is undoubtedly correct, but scarcely deserves to rank as an explanation, if by explanation is meant that anyone who knew that iron has a high permeability should have foreseen the result. The treatment in the paper is based on the discussion of the problem given by Moullin.* It is assumed that, in order for there to be a certain flux density B at a point in the iron, there will also have to be a magnetic force H at that point. The problem of flux propagation, therefore, becomes the problem of the propagation of magnetic force. If the contributions of pole strength within the material are neglected, it is clear that the surface polarity on the iron must provide the mechanism by which the iron can maintain a substantially constant magnetic force throughout a particular iron circuit. If there is to be surface polarity on the iron, there will also be flux emerging from the iron. This flux is commonly called leakage flux and is often regarded as an imperfection in the design of the apparatus. But if we consider an iron-cored transformer in which the primary and secondary windings are wound on different portions of the core, it follows that it is the leakage flux that gives rise to the mutual flux. So far from being an imperfection in the design, the leakage flux is merely a manifestation of the surface polarity by which the iron maintains a constant mutual flux around the core. Without leakage flux there would be hardly any mutual flux in such a transformer. If the permeability of the iron is very large, a small surface polarity will be sufficient to maintain the flux constant around the iron. The fractional value of leakage flux will therefore tend to zero when the permeability tends to infinity, but it is clear that the absolute value of the leakage flux will not tend to zero. The problem should be considered with reference to Fig. 1.

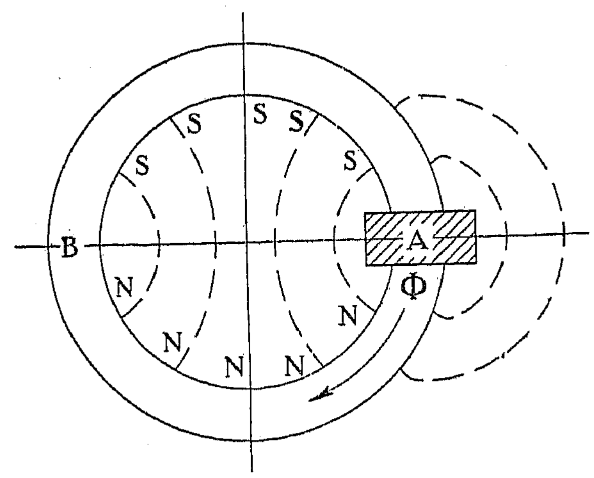


Fig. 1

In order to test the foregoing hypothesis, it was decided to derive a mathematical solution for the leakage flux in a certain problem and then to test the result experimentally. To simplify the mathematics it was decided to restrict the problem to two dimensions and to assume that the iron had constant permeability. The problem chosen was that of an infinitely long cylinder magnetized by a single current loop (see Figs. 2 and 3). The leakage flux was defined as the difference between the fluxes crossing sections A and B of the cylinder, respectively. The ratio of the fluxes at A and B was calculated to be

$$\frac{\Phi_A}{\Phi_B} = \frac{\mu \log \frac{a}{b} - 2 \log \left(1 - \frac{c}{b}\right) - 2 \log \left(1 - \frac{a}{d}\right)}{\mu \log \frac{a}{b} - 2 \log \left(1 + \frac{c}{b}\right) - 2 \log \left(1 + \frac{a}{d}\right)}$$

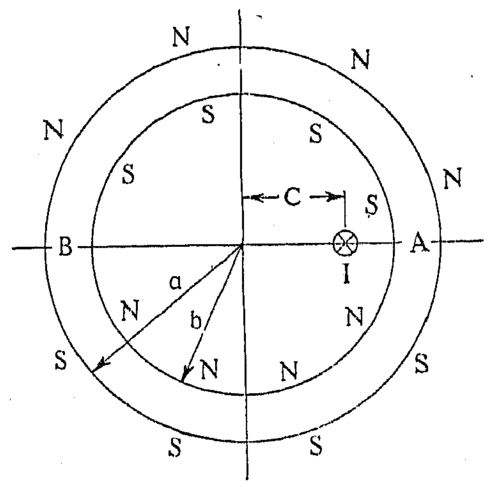


Fig. 2

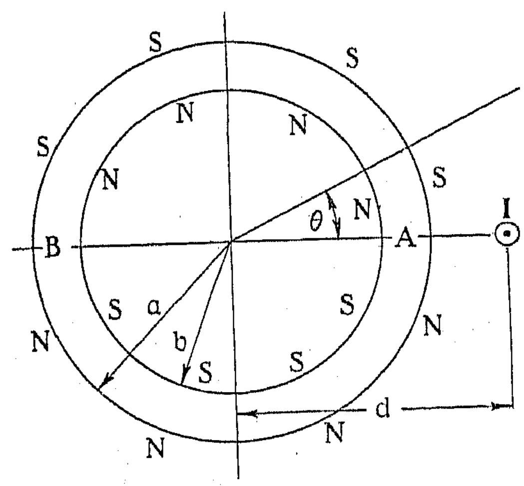


Fig. 3

Fig. 4 shows the percentage leakage flux

$$\frac{\Phi_A - \Phi_B}{\Phi_A} \times 100$$

calculated for the case $a/b = 3/2$ and $c/b = a/d$ at different permeabilities and positions of the magnetizing loop. The curves show that the leakage flux increases considerably as the current is placed closer to the wall of the cylinder. A magnetizing winding consisting of a single loop is of course an extreme case, and in power transformers the leakage flux may be of little importance. But transformers are required in certain types of calculating machine, and in such applications it may be desirable to estimate the order of the discrepancy between voltage ratio and turns ratio.

It may well be doubted whether results based on the assump-

* MOULLIN, E. B.: "Principles of Electromagnetism" (Oxford University Press), Third edition, pp. 164–168.

Mr. Hammond is in the Department of Engineering, University of Cambridge.

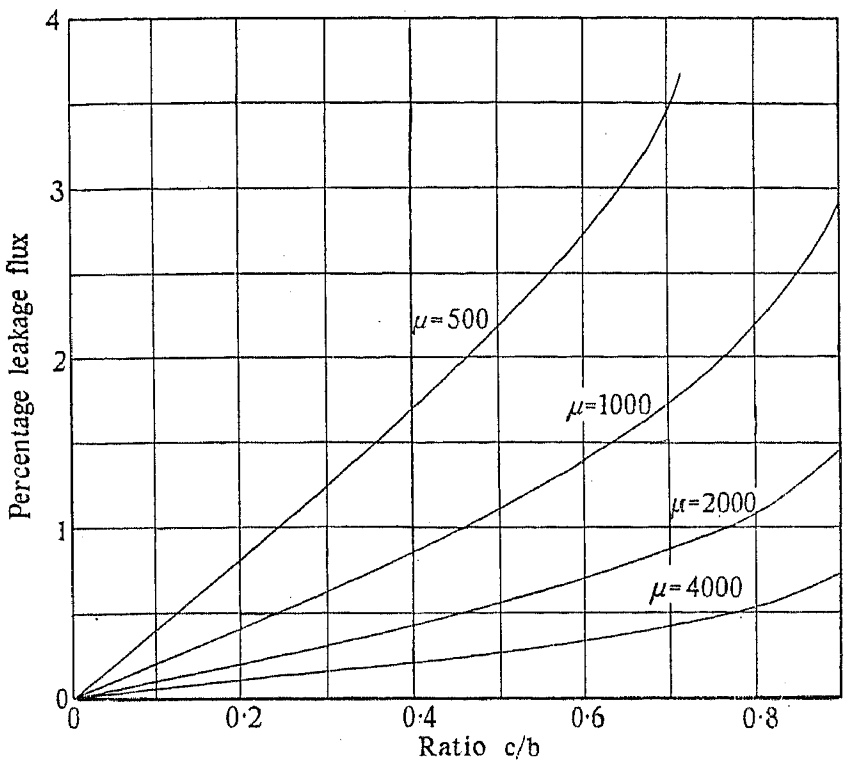


Fig. 4.—Theoretical leakage flux.

tion of constant permeability of the material can be of any value when iron is used, especially if there is considerable saturation. The experimental results show that the results do, in fact, apply to iron of widely different permeabilities. The reason for this remarkable fact is that the actual value of the permeability does not matter greatly, so long as it is much larger than unity. In the paper it is shown that the boundary condition at the surface of the cylinder is

$$2\pi\sigma = \frac{\mu - 1}{\mu + 1} H_r$$

where σ is the local density of polarity and H_r is the magnetic force normal to the surface at the point considered. It will therefore be seen that the surface polarity is to a large extent independent of the permeability, and thus the absolute leakage flux is also independent of the permeability.

The experimental work was carried out on iron tubes of varying length built up from silicon-iron stampings lightly insulated on one side. The tubes were mounted, with their axes horizontal, on a slide which could be moved transversely to a long copper rod which carried the 50c/s magnetizing current. Search coils were wound on the tubes at opposite ends of a diameter and could be connected either in series or in opposition. An "integrating circuit" was used in order to obtain traces of flux density on an oscillograph. Fig. 5 shows some typical traces, and it will be seen that the leakage flux [Fig. 5(*b*)] is practically independent of the permeability, although there is considerable saturation [see Figs. 5(*c*) and 5(*e*)]. Fig. 6 shows the percentage leakage flux with different magnetizing currents and also the calculated curve for constant permeability. In Fig. 7 the shape of the calculated and experimental leakage-flux curves is compared on the basis of choosing a value for the permeability which will make the theoretical and experimental curves coincide for the position of the magnetizing current when $c/b = 0{\cdot}65$ (see Fig. 2). The "equivalent permeabilities" obtained in this way are plotted in Fig. 8 for tubes of different lengths. The permeabilities obtained from a d.c. reversal curve are plotted in Fig. 8 for comparison. Some of the tubes had "guard rings" in the form of further stampings placed at their ends in order to simulate more closely the case of an infinitely long tube. It is seen from Fig. 8 that there is good experimental agreement with the calculation, and that any discrepancy arises chiefly from the short length of the tubes.

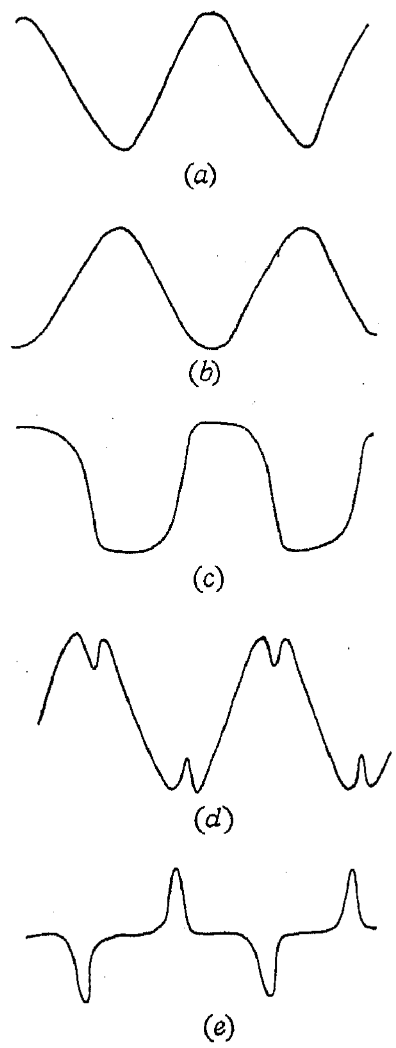


Fig. 5.—Waveforms.

(*a*) Current.
(*b*) Leakage flux.
(*c*) Mutual flux.
(*d*) "Leakage" voltage.
(*e*) "Mutual" voltage.

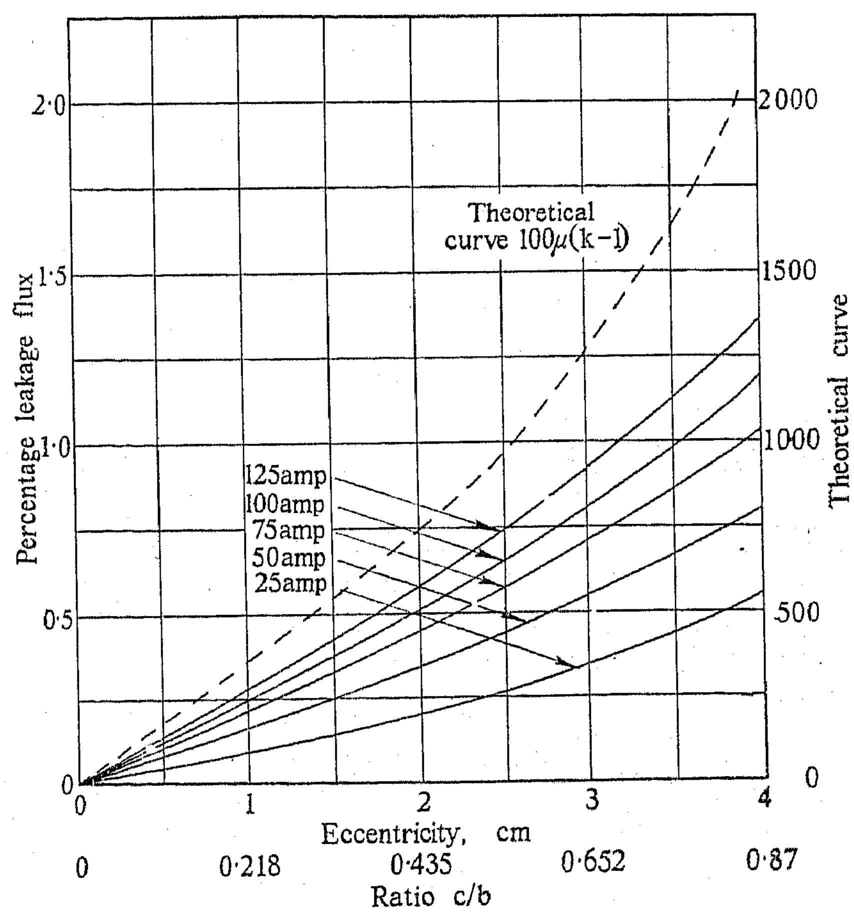


Fig. 6.—Percentage leakage flux in 35 cm tube.

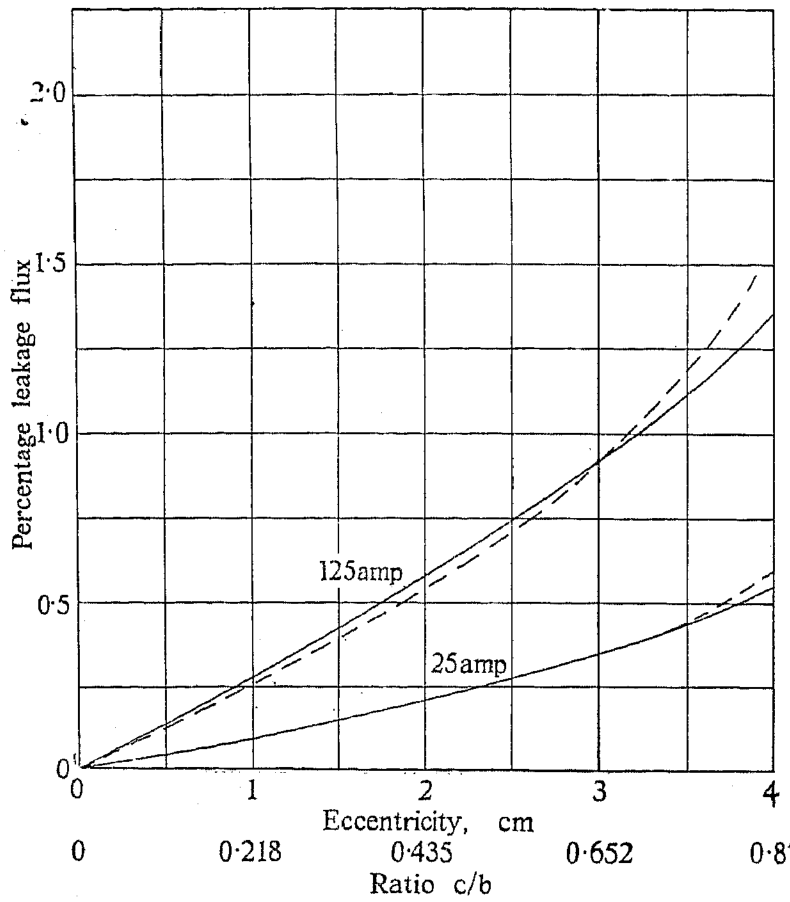


Fig. 7.—Comparison of shape of theoretical and experimental leakage-flux curves.

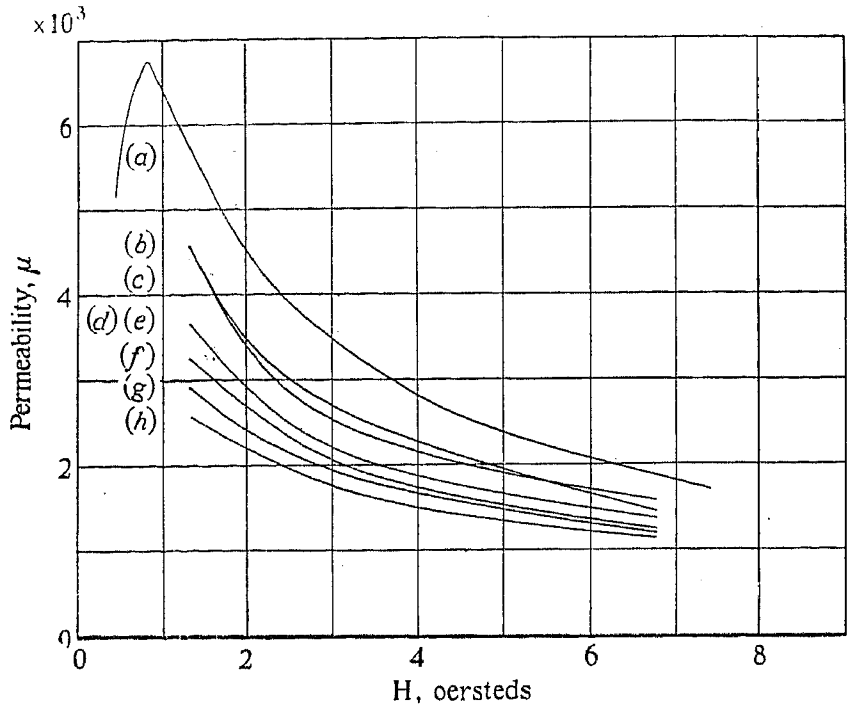


Fig. 8.—Comparison of equivalent permeabilities.

(a) D.C. reversal curve.
(b) 8·75 cm specimen packed both sides.
(c) 17·5 cm specimen packed both sides.
(d) 35 cm specimen.
(e) 17·5 cm specimen packed one side.
(f) 8·75 cm specimen packed one side.
(g) 17·5 cm specimen.
(h) 8·75 cm specimen.

It has been shown that surface polarity (or leakage flux) gives to the iron its remarkable property of being able to maintain a constant flux around an iron circuit magnetized by a concentrated coil. Good experimental agreement is obtained with calculations, even when these are based on the apparently drastic assumption of constant permeability of the iron. It therefore becomes possible to estimate the leakage flux in any particular set of iron-ring stampings and to estimate the order of the flux leakage in more complicated iron circuits.

SIGNAL/NOISE PERFORMANCE OF MULTIPLIER (OR CORRELATION) AND ADDITION (OR INTEGRATING) TYPES OF DETECTOR

621.376.23 : 621.38 Monograph No. 120 R

D. G. TUCKER, D.Sc., Ph.D., Member

(Digest *of a paper published in February,* 1955, *as an* Institution Monograph *and to be republished in Part C of the* Proceedings.)

The signal/noise performance of a detector may often be improved by multiplying or adding together two or more "scans" in which the signals coincide in time or are correlated, and in which the noise backgrounds are uncorrelated. The adding process is quite familiar and occurs in radar systems by virtue of the afterglow of a cathode-ray-tube screen; it can also be arranged by means of storage tubes or magnetic-drum or -tape systems, and there is no theoretical limit to the number of additions. In practice, the signal generally moves in the time scale, and so there is a practical limit to the number of scans which can be added with effective coincidence of signal position. The adding process is often referred to as integration—or pulse-to-pulse integration, or scan-to-scan integration—by analogy with the similar results of filtration, which is a true integration process.

Multiplying scans together is a less familiar process, although it occurs in the interferometer techniques used in radio-astronomy[1,2] and elsewhere. In such techniques, the "scans" which are multiplied are those obtained simultaneously on two separate arrays. But obviously successive scans of a single array may be multiplied together if a storage system is used. The multiplying-type detector is often referred to as a correlation detector, because its d.c. output represents the correlation coefficient (or function) of the two input signals.

Multiplying or adding may be performed on single-array systems without storage devices if two or more independent scans can be simultaneously obtained by using, for example, different frequency-bands. Thus it can be seen that the scope for multiplier or addition detectors is quite wide.

Multiplying or adding may generally be done on the incoming signal or after rectification, although when different frequency-bands are involved, as in the previous paragraph, there is no choice but to process the signals after rectification. If the processing is done on the incoming signal, attention must be paid to phasing of the signal.

The purpose of this paper is to examine the signal/noise performance of these processes. No other aspect is considered in detail. The analysis is based on a simplified representation of noise by a series of small but equal-amplitude tones of different frequencies, and lack of correlation between two noise waveforms is shown by taking different series, each with a quite independent set of frequencies. This representation, while possibly not rigorous, is thought to give no error in the present

Dr. Tucker is in the Royal Naval Scientific Service.

application. The signal/noise criteria are those defined in a previous paper,[3] and take no particular account of special probability distributions of the noise. As the author is concerned mainly with narrow-band pulse systems—i.e. those with optimum input filtration—in which there is little scope for post-detector integration (i.e. low-pass filtration), there is no discussion of the effect of such integration. However, it is throughout assumed that the h.f. components (e.g. the input frequency-band upwards) are removed from the detector outputs, immediately after rectifying.

The general conclusions reached are that multiplying is not, in general, more effective than adding, and being more difficult has often, therefore, little to recommend it. It does have a slight advantage due to its square-law dynamic response, but this can more simply be obtained by using square-law rectifiers. When done on the unrectified signal, multiplying has the big advantage of giving a final output which contains no d.c. component in the absence of correlated signal. This is very important when the background is liable to vary, since then, in a system using addition followed by rectification, the fluctuating d.c. background may be difficult to bias out, and will often more than offset the advantage in signal/noise performance this system has at low input-signal/noise ratios; but the multiplier, having no d.c. background, will be free from such difficulties. This is the main advantage of the multiplier over the addition system in radio-astronomy. Figs. 1(*a*)–1(*c*) are oscillograms of typical background (white noise) waveforms at the output of a linear detector, square-law detector, and multiplier (multiplying two uncorrelated noise inputs without prior rectification). The zero voltage axis is shown in all three waveforms, and the absence of a d.c. output in the multiplier is clear. For completeness, Fig. 1(*d*) shows the multiplier output when the noise inputs are correlated exactly; a d.c. output now exists.

The analysis is in terms of one addition or one multiplication, but evidently the processes can be repeated as long as the signal pulses can be added in phase or in coincidence, and the end results are easily deduced from the analysis and Tables presented here.

Results

Details of the analysis of the signal/noise performance of the various detector arrangements (except for those previously published[3,4]) are given in the appendices to the paper, and numerical results over the range of R_1 (input-signal/noise ratio) from 0·1 up to 4 are given in Tables 1 and 2.

It should be noted that some of the arrangements analysed give dynamic characteristics which are, at least nominally,* linear; others are basically quadratic, while one or two have a quartic response. Table 1 includes those cases which are nominally linear, and Table 2 those which are quadratic. The cases of quartic-type response (e.g. multiplication of the outputs of square-law rectifiers) are not worked out numerically, since they are not thought to be of very great interest; but their analysis

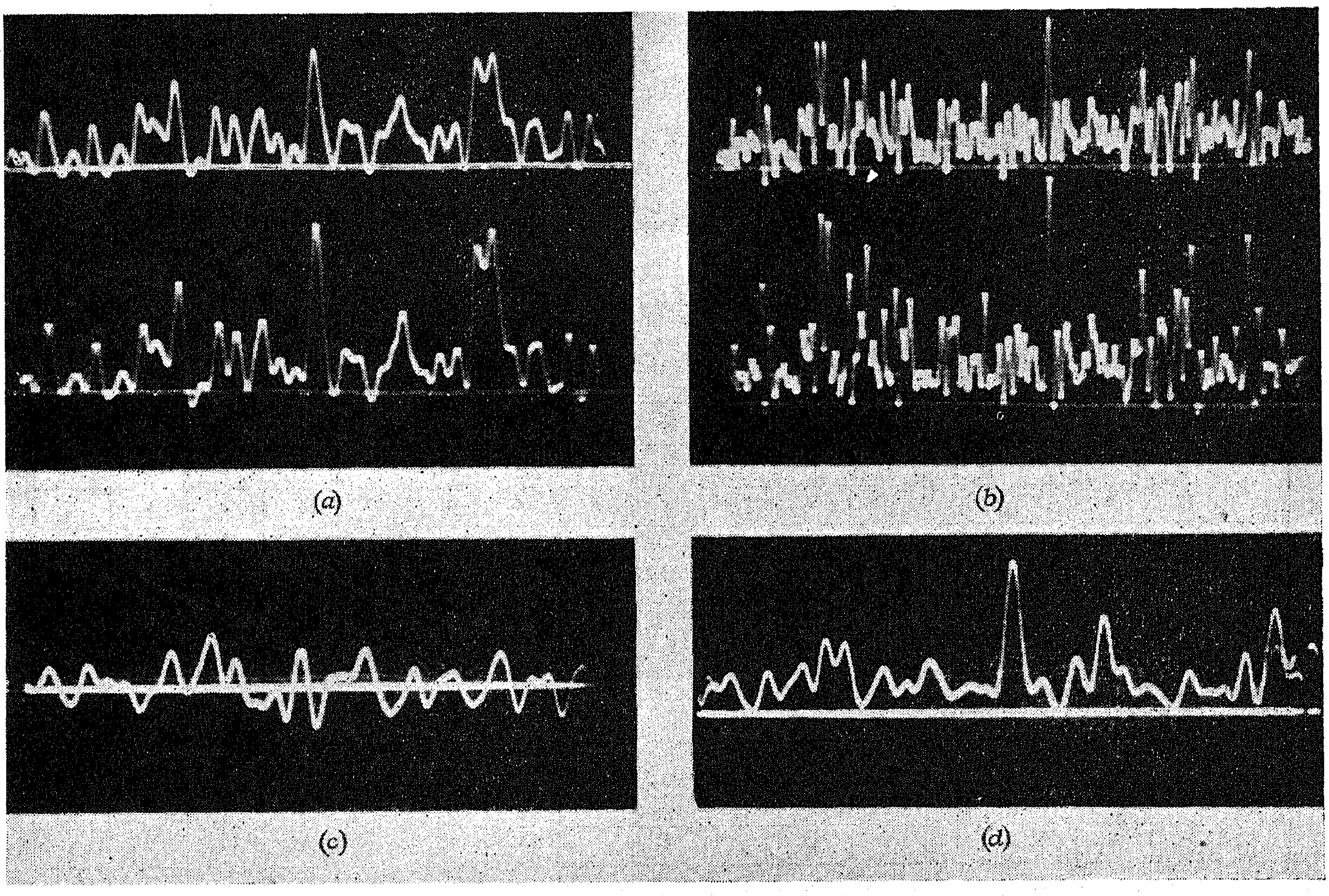


Fig. 1.—Oscillograms of noise waveforms from various detectors.

Outputs are smoothed to leave only envelope frequencies.

(*a*) Outputs of linear (top) and square-law (bottom) detectors for the same input waveform.
(*b*) As (*a*), but with slower time-scan.
(*c*) Output of multiplier (multiplying unrectified signals) when the two noise inputs are uncorrelated.
(*d*) As (*c*), but when noise inputs are completely correlated (i.e. same waveform applied to both inputs).

* "Nominally linear" because at low signal/noise ratios the behaviour of the linear rectifier is effectively quadratic.

TABLE 1

Detectors with Linear Dynamic Response

R_1	R_{A1}							R_{A2}						
	A	B	C	D	A	B	C	A	B	C	D	A	B	C
	X	X	X	X	Y	Y	Y	X	X	X	X	Y	Y	Y
4	3·53	4·74	3·75	6·66	4·12	5·75	4·87	4·12	5·75	4·39	5·75	4·12	5·75	4·55
2	2·08	2·68	2·15	3·83	2·24	3·0	2·68	2·24	3·0	2·34	3·0	2·24	3·0	2·49
1	1·39	1·62	1·42	2·41	1·41	1·73	1·63	1·41	1·73	1·45	1·73	1·41	1·73	1·54
0·5	1·12	1·20	1·12	1·70	1·12	1·22	1·20	1·12	1·22	1·13	1·22	1·12	1·22	1·16
0·2	1·02	1·04	1·02	1·28	1·02	1·04	1·04	1·02	1·04	1·02	1·04	1·02	1·04	1·027
0·1	1·005	1·01	1·005	1·14	1·005	1·01	1·01	1·005	1·01	1·005	1·01	1·005	1·01	1·007

R_1	R_{B1}							R_{B2}						
	A	B	C	D	A	B	C	A	B	C	D	A	B	C
	X	X	X	X	Y	Y	Y	X	X	X	X	Y	Y	Y
4	4·5	6·8	6·42	5·65	1·52	1·65	1·22	6·9	10·4	9·69	5·65	6·24	9·5	8·44
2	1·85	2·9	2·59	2·82	1·11	1·33	0·86	2·7	4·4	3·8	2·82	2·48	4·0	3·34
1	0·65	1·1	0·90	1·41	0·59	0·85	0·50	0·85	1·55	1·21	1·41	0·83	1·46	1·12
0·5	0·22	0·4	0·29	0·71	0·214	0·36	0·23	0·22	0·44	0·32	0·71	0·24	0·44	0·32
0·2	0·039	0·078	0·053	0·282	0·04	0·08	0·05	0·039	0·078	0·054	0·282	0·04	0·08	0·054
0·1	0·010	0·02	0·013	0·141	0·01	0·02	0·013	0·010	0·02	0·014	0·141	0·01	0·02	0·013

A = Linear rectifier.
B = Addition followed by linear rectifier.
C = Addition after linear rectifiers.
D = Coherent detector.

X for coherent signal.
Y for noise signal.

TABLE 2

Detectors with Quadratic Dynamic Response

R_1	R_{A1}								R_{A2}							
	E	F	G	H	E	F	G	H	E	F	G	H	E	F	G	H
	X	X	X	X	Y	Y	Y	Y	X	X	X	X	Y	Y	Y	Y
4	11·37	28·4	14·6	12·34	17·0	45·8	22·1	21·69	12·7	28·3	16·84	14·28	17·0	32·5	22·0	21·47
2	4·00	8·63	4·72	4·17	5·0	12·0	6·09	6·52	4·12	6·4	5·00	4·44	5·0	8·55	6·07	6·45
1	1·87	3·13	1·98	1·90	2·0	3·64	2·17	2·49	1·87	2·22	2·00	1·91	2·0	2·63	2·17	2·46
0·5	1·24	1·58	1·25	1·24	1·25	1·62	1·27	1·39	1·24	1·275	1·25	1·24	1·25	1·32	1·27	1·37
0·2	1·04	1·095	1·04	1·04	1·04	1·10	1·04	1·06	1·04	1·04	1·04	1·04	1·04	1·045	1·038	1·06
0·1	1·01	1·025	1·01	1·01	1·01	1·025	1·01	1·01	1·01	1·01	1·01	1·01	1·01	1·01	1·010	1·015

R_1	R_{B1}								R_{B2}							
	E	F	G	H	E	F	G	H	E	F	G	H	E	F	G	H
	X	X	X	X	Y	Y	Y	Y	X	X	X	X	Y	Y	Y	Y
4	2·8	3·94	3·88	3·94	0·94	0·97	1·09	0·80	16·0	22·6	25·33	22·6	16·0	22·6	25·5	22·6
2	1·33	1·89	1·79	1·89	0·80	0·88	0·97	0·64	4·0	5·66	6·05	5·66	4·0	5·66	6·18	5·66
1	0·60	0·80	0·76	0·80	0·50	0·635	0·65	0·44	1·0	1·42	1·39	1·42	1·0	1·41	1·44	1·414
0·5	0·25	0·29	0·26	0·29	0·20	0·278	0·26	0·22	0·25	0·35	0·319	0·35	0·25	0·354	0·33	0·354
0·2	0·04	0·054	0·049	0·054	0·038	0·054	0·05	0·052	0·04	0·057	0·051	0·057	0·04	0·057	0·052	0·057
0·1	0·01	0·014	0·013	0·014	0·010	0·014	0·013	0·014	0·01	0·014	0·013	0·014	0·01	0·014	0·013	0·014

E = Square-law rectifier.
F = Multiplier.
G = Multiplier after linear rectifiers.
H = Addition after square-law rectifiers.

X for coherent signal.
Y for noise signal.

is included in the appendices, so that all basic information is available should numerical results ever be required.

The criteria of signal/noise performance used are those defined in a previous paper,[3] namely

$$R_{A1} = \frac{\text{(d.c. output due to signal and noise)} + \text{(r.m.s. output of l.f. noise when signal present)}}{\text{d.c.} + \text{r.m.s. value of l.f. output when signal absent.}}$$

$$R_{A2} = \frac{\text{Root mean square of d.c. plus l.f. components when signal present}}{\text{Root mean square of d.c. plus l.f. components when signal absent}}$$

$$R_{B1} = \frac{\text{Change in direct current on application of signal}}{\text{l.f. noise when signal present}}$$

$$R_{B2} = \frac{\text{Change in direct current on application of signal}}{\text{l.f. noise when signal absent}}$$

The coherent detector[4] has been included among the linear detectors (it is obviously applicable only to coherent-tone signals and not to noise signals) for comparison. The outputs of two coherent detectors with uncorrelated backgrounds can evidently be added together or multiplied just as with other detectors, and in the case of the R_B criteria, an addition gives an improvement of $\sqrt{2}$ (i.e. 3 dB) just as with linear or square-law rectifiers. But these cases are not included in the Tables.

In Table 1 it can be observed that whichever criterion is used, adding before rectifying is generally better than adding after rectifying. On criteria R_{B1} and R_{B2} the improvement relative to a single rectifier is a factor of about $\sqrt{2}$ at high signal/noise ratios; but at low signal/noise ratios, adding before rectifying profits from the square-law nature of the linear-detector response at low ratios and doubles the values of R_{B1} and R_{B2}, while adding after rectification merely continues to give the $\sqrt{2}$ improvement.

In Table 2 it can be observed that:

(*a*) On criteria R_{A1} and R_{A2}, multiplying the input signals together is far better than multiplying after rectification—in fact, the latter is little better than a single output.

(*b*) On criteria R_{B1} and R_{B2}, there is little to choose between multiplying the input signals together, multiplying after rectification, or adding after square-law rectifiers—except R_{B1} with a noise-signal, which is a case of no practical significance, as R_{B1} is not a suitable criterion for noise signals—but all three cases are about $\sqrt{2}$ times better than a single square-law rectifier. Clearly, adding before square-law rectification gives a doubling of the values of R_{B1} and R_{B2} (this case is not tabulated, being obvious) and is therefore superior to all other cases.

The significance of a comparison between Tables 1 and 2 is not immediately clear, and one aspect of it is discussed in the next Section.

Linear versus Quadratic Dynamic Response

It is clear from Tables 1 and 2 that those detection circuits with a quadratic dynamic response have considerably higher values of R_{A1}, R_{A2} and R_{B2} at the upper ranges of R_1, and noticeably higher values at the lower ranges, than those circuits with a linear response. It is tempting to think that this means that better detection is obtained by using circuits with quadratic (or higher-order) response instead of linear circuits. In a previous paper[3] doubt was cast on the truth of this conclusion, and the relationship between subjective probability of detection of a signal against noise and the objective signal/noise ratio and its derived criteria (such as R_A and R_B) has never been clear. There can, of course, be no unique relationship between probability of detection and signal/noise criteria, because subjective detection is a function of many factors, such as display efficiency, operator fatigue, environment, any pattern or shape in the signal as seen on the display, assistance by aural presentation, etc. The real question is whether, for any given set of display and operating conditions, signal/noise criteria can be used to compare the merits of different kinds of detector circuit. They are useful criteria only if they can be so used.

Recently, subjective measurements of threshold-detection performance have been made, and a limited measure of support for the use of signal/noise criteria has been obtained. So far as intensity-modulated displays are concerned, it has been shown —theoretically[8] as well as practically—that contrast, obtained by the use of bias, leads to better detection when the dynamic range of the display, measured in terms of "just-noticeable differences," is limited; criterion R_A (i.e. either R_{A1} or R_{A2}), which has been considered appropriate for intensity-modulated displays, does in fact give higher values when contrast is applied, and to this extent may be considered a reliable criterion. But the problem of the performance of intensity-modulated displays is very complicated and no comprehensive data are yet available.

The problem of detection by A-scans is much easier, and comparative results of linear and square-law detectors using an A-scan display are available. Experimental results obtained for probability of detection of a pulse of 20 millisec duration against "white" noise restricted to a 200 c/s bandwidth, using A-scan cathode-ray-tube display with a scan duration of 250 millisec, and averaged over several hundred observations by several different observers, show no large difference in detectability between single linear and square-law detectors. Fig. 2 shows a

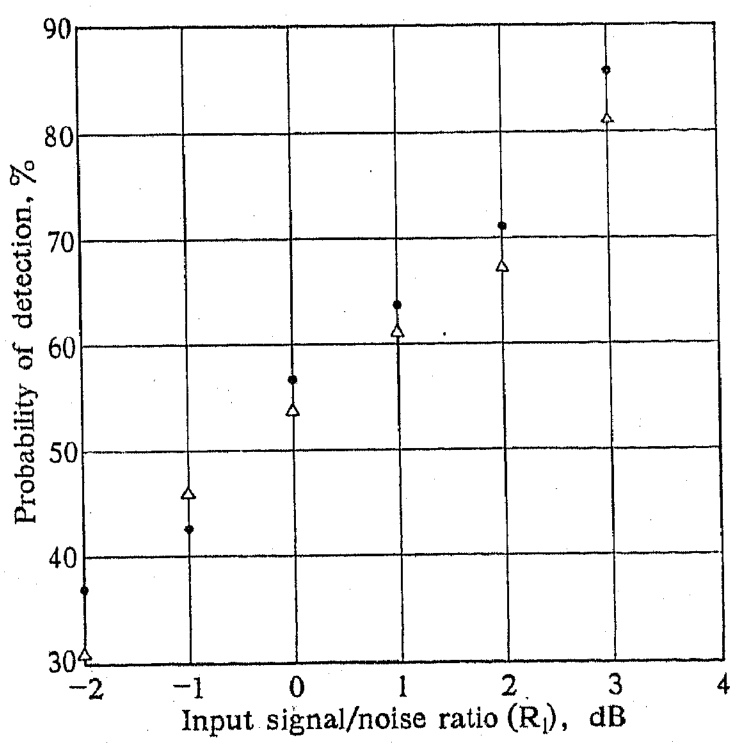


Fig. 2.—Subjective detection results for three operators.
● Square-law detector.
△ Linear detector.

graph of percentage probability of detection against input signal/noise ratio (R_1) for the average of a hundred observations at each value of R_1 by each of three observers on each type of detector, using for each observation a single scan on a long-persistence cathode-ray tube. The pulse, which was of coherent tone, was applied in one of ten positions on the scan, and the observer had to state in which position the pulse appeared. No "nil" record was allowed, and the percentage probability of detection is the actual percentage of correct detections obtained. Results are shown for linear and square-law detectors. It will be

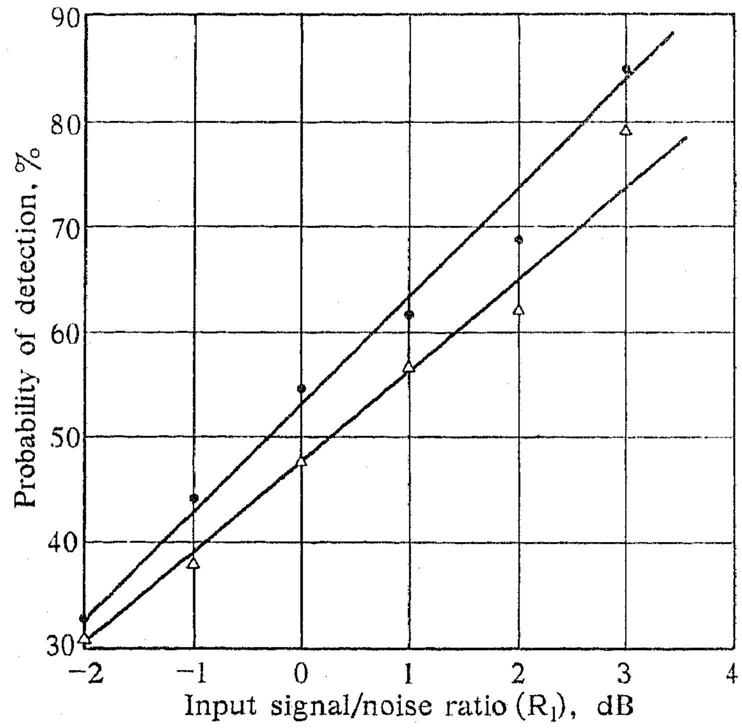


Fig. 3.—Subjective detection results for one more-consistent operator.
● Square-law detector.
△ Linear detector.

seen that over the range of R_1 from -2 to $+3$ dB there is only a slight difference in favour of the square-law detector. Fig. 3 shows a similar pair of curves based on one hundred observations for each value of R_1 on each type of detector, made by one particular observer (among the three) whose results were more consistent than those of the two other observers; here the difference in favour of the square-law detector is greater. Assuming these results are, in fact, more reliable than the others, and that the straight-line approximations are valid, we shall

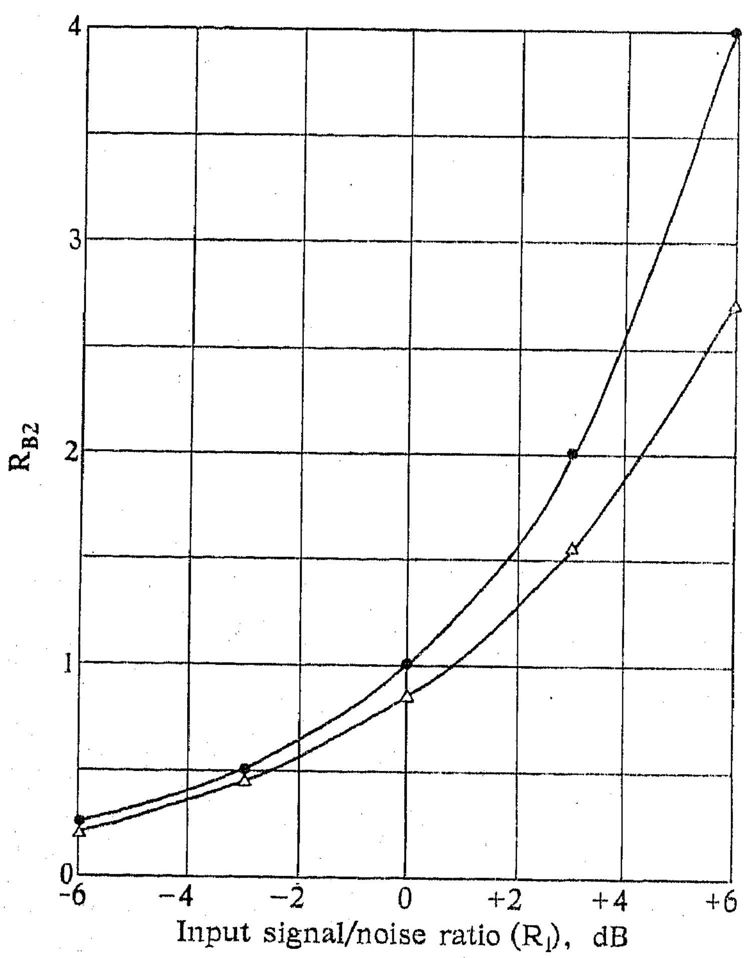


Fig. 4.—Graphs of detection criterion R_{B2} against input signal/noise ratio, R_1.
● Square-law detector.
△ Linear detector.

now try to relate these probabilities of detection to the criterion R_{B2} (which is considered the appropriate one for *A*-scans with short pulses) using the calculations summarized in Tables 1 and 2.

Fig. 4 shows a graph of R_{B2} against R_1 for both linear and square-law detectors. From this and Fig. 3, points can be plotted, as shown in Fig. 5, relating probability of detection to R_{B2}. It will be seen that over the range of probabilities from 30 to 85%, the points lie remarkably closely on one and the same curve, as shown. If the straight-line approximations of Fig, 3 are not used, but instead the actual observed values are taken in constructing Fig. 5, then the points lie equally well on

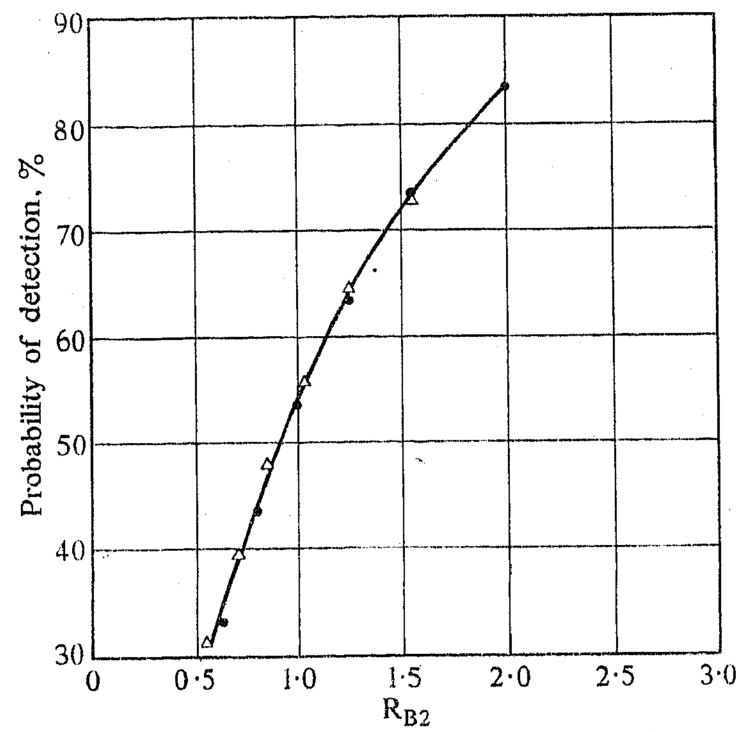


Fig. 5.—Graphs of probability of detection against R_{B2} for the one more-consistent operator.
● Square-law detector.
△ Linear detector.

a single curve except for those corresponding to $R_1 = +3$ dB. These results support the hypothesis that for this particular kind of display the relationship between probability of detection and R_{B2} is independent of the type of detector circuit used, so that the signal/noise criterion may well be perfectly reliable as a means of comparing the performance of different detector circuits. There is, of course, no proof as yet, and the above comparisons have been made only on the simplest cases. Moreover, the data used are not very comprehensive, and are selected as being the most reliable among a larger batch of data, which, taken as a whole, give rather less impressive results. The subjective investigation of detectability is proceeding.

The matter most open to doubt is whether the divergence of values of R_A and R_B as between linear and square-law responses at large R_1 represents any practical difference in detectability. At these larger values of R_1 the probability of detection is very high, and consequently it cannot be altered very much by changing from linear to square-law detectors. But it is possible, of course, and even probable, that the greater contrast of the display obtained with the square-law response gives the operator more confidence in his detections, and—partly because of this—reduces operator fatigue. This would lead eventually to better detection under operational conditions. There is, however, no experimental evidence available on this aspect of the problem.

REFERENCES

(1) RYLE, M.: "A New Radio Interferometer and its Application to the Observation of Weak Radio Stars," *Proceedings of the Royal Society*, A, 1952, **211**, p. 351.

(2) HANBURY BROWN, R., JENNISON, R. C., and DAS GUPTA, M. K.: "Apparent Angular Sizes of Discrete Radio Sources," *Nature*, 1952, **170**, p. 1061.

(3) TUCKER, D. G., and GRIFFITHS, J. W. R.: "Detection of Pulse Signals in Noise," *Wireless Engineer*, 1953, **30**, p. 264.

(4) TUCKER, D. G.: "The Synchrodyne and Coherent Detectors," *ibid.*, 1952, **29**, p. 184.

(5) BENNETT, W. R.: "Response of a Linear Rectifier to Signal and Noise," *Journal of the Acoustical Society of America*, 1944, **15**, p. 164.

(6) JAHNKE, E., and EMDE, F.: "Tables of Functions" (Leipzig, 1938, or New York, 1943).

(7) TUCKER, D. G.: "Linear-Rectifiers and Limiters: Effect of Complex Signals and Noise," *Wireless Engineer*, 1952, **29**, p. 128.

(8) TUCKER, D. G., and GRIFFITHS, J. W. R.: "Detection of Pulse Signals in Noise by Intensity-Modulated Visual Display: Theory in Terms of Just-Noticeable Differences," *ibid.* (in course of publication).

A DIRECT-READING WAVEGUIDE STANDING-WAVE DETECTOR FOR USE AT LOW POWER LEVELS

621.317.742 : 621.372.8 Monograph No. 119 R

H. V. SHURMER, M.Sc., Ph.D., Associate Member

(DIGEST *of a paper published in January, 1955, as an* INSTITUTION MONOGRAPH *and to be republished in Part C of the* PROCEEDINGS.)

Voltage standing-wave ratios are measured on a direct-reading instrument by sampling the reflected power with a directional coupler. Ratios of up to 0·90 are measured to an accuracy of better than ±0·01 for power levels at the load of down to $5\,\mu$W.

The r.f. source is square-wave modulated, and the signal appearing at the detector arm of the coupler is fed into a frequency-selective amplifier. The wide range of input voltages corresponding to the spread of v.s.w.r.'s is accommodated by the use of a stepped attenuator built into the amplifier, four steps of attenuation being required.

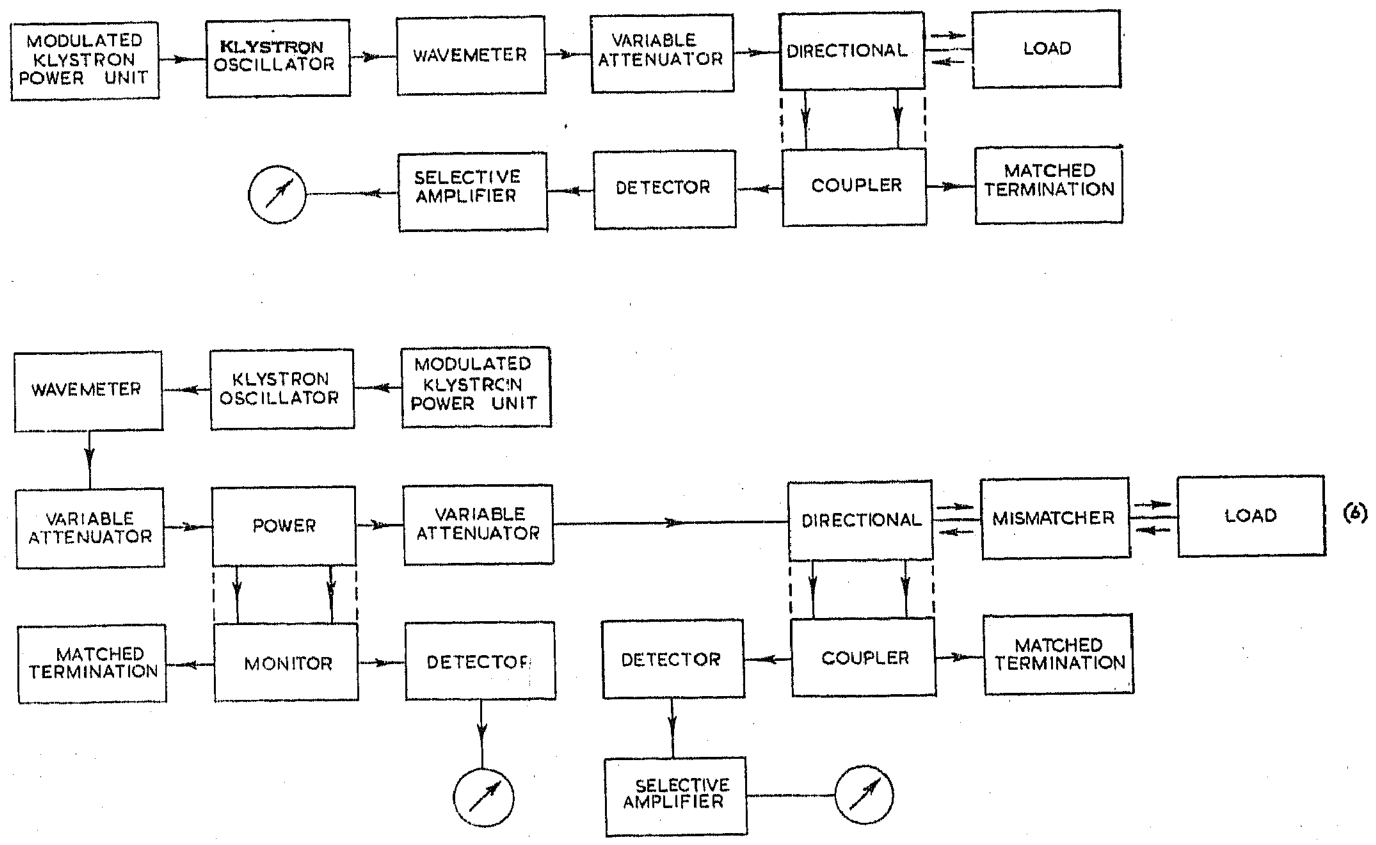


Fig. 1.—Block schematic of direct-reading v.s.w.r. equipment.
(*a*) Basic details. (*b*) Complete equipment.

The low power level of $5\,\mu$W is required for testing detector crystal valves. At this power level tight coupling is required to the detector arm, in which case it is difficult to obtain the necessary directivity. To permit the use of a relatively poor coupler, the wave in the detector arm due to the forward travelling wave in the main waveguide and also due to imperfect termination at the fourth arm of the coupler are cancelled by a preset mismatcher situated in front of the load.

The arrangement of components used in the equipment is shown in Fig. 1.

Dr. Shurmer is with the British Thomson-Houston Co., Ltd.

REFERENCES

(1) RATCLIFFE, P. M.: "The Directional Coupler as an Impedance Measuring Device," *Marconi Instrumentation*, 1953, **4**, p. 9.

(2) SMITH, P. H.: "Transmission Line Calculator," *Electronics*, 1939, **12**, p. 29.

(3) SMITH, P. H.: "An Improved Transmission Line Calculator," *ibid.*, 1944, **17**, p. 130.

RESONANT-CAVITY MEASUREMENTS OF THE RELATIVE PERMITTIVITY OF A D.C. DISCHARGE

621.317.335.3 : 621.372.413 : 621.3.015.3 Monograph No. 122 M

K. W. H. FOULDS, Ph.D., B.Sc.(Eng.), Graduate

(DIGEST *of a paper published in March,* 1955, *as an* INSTITUTION MONOGRAPH *and to be republished in Part C of the* PROCEEDINGS.)

LIST OF SYMBOLS

e, m = Charge and mass of an electron, respectively.
f = General symbol for the impressed frequency.
f_0 = Resonant frequency of the cavity.
δf = Change in resonant frequency of the cavity caused by some discharge current.
I_a = Current through the discharge tube.
N = General symbol for electron density.
Q = Q-factor $= \dfrac{2\pi f \times \text{Energy stored}}{\text{Power loss}}$
Q_0 = Q-factor of the cavity with no discharge current.
Q_L = Q-factor of the cavity for some specified discharge current.
ϵ = General symbol for permittivity.
ϵ_0 = Permittivity of free space.
ϵ_1' = Relative permittivity of the discharge.
θ_0 = Height of the resonance curve for no discharge current.
θ_L = Height of the resonance curve for some specified discharge current.
ν = Electron collisional frequency.
σ = Conductivity of the discharge.
$\omega = 2\pi f$.

The relative permittivity of an ionized gas and its conductivity are given by the equations

$$\epsilon_1' = 1 - \frac{Ne^2}{m(\omega^2 + \nu^2)}\frac{1}{\epsilon_0} \quad . \quad . \quad . \quad . \quad (1)$$

$$\sigma = \frac{Ne^2\nu}{m(\omega^2 + \nu^2)} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If the mean free path of the electrons in the discharge is large ν is very much smaller than ω at microwave frequencies, and so far as an impressed radio-frequency electromagnetic wave is concerned, the conductivity of the ionized gas is zero, and the relative permittivity takes the form

$$\epsilon_1' = 1 - \frac{Ne^2}{m\omega^2\epsilon_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Many attempts to verify these equations have been made,[4–12] based on experiments in which a gaseous discharge forms a dielectric of variable permittivity between two condenser plates. It has been shown[6] that the anomalous results obtained in the early experiments were due to the formation of positive-ion sheaths around the condenser plates. In recent years new techniques have been evolved for measuring the dielectric properties of materials in the microwave band of frequencies.[13] The present paper describes experiments in which the relative permittivity of a d.c. discharge has been measured by using these now well-established techniques. The principle underlying all the experiments is as follows. The gaseous discharge forms part of a resonant circuit, which at very high frequencies takes the form of a cavity resonator, and the relative permittivity of the discharge, for a particular discharge current, is calculated from the observed change in the resonant frequency of the system caused by that current. Similarly, the conductivity can be calculated from the observed change in the Q-factor of the resonant system.

Fig. 1 shows the discharge tube lying along the axis of a cylindrical cavity which resonates in the H_{011} mode. This mode was chosen so that the single cavity would resonate conveniently over the frequency range 2 000–3 000 Mc/s. The frequencies at which the measurements were made were chosen carefully so

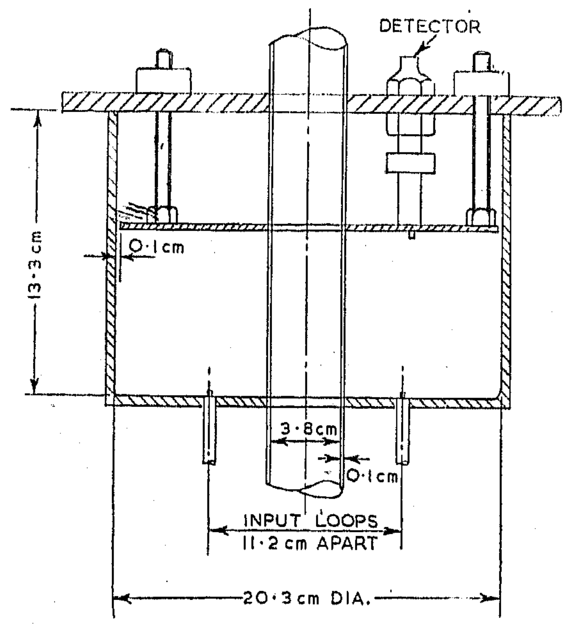


Fig. 1.—The resonant cavity.

that the cavity resonated in a sensibly undisturbed H_{011} mode, and this was verified experimentally. Two different discharges have been studied. The initial and rather exploratory experiments were carried out using a commercial 80-watt fluorescent lighting tube maintained by a d.c. supply through suitable ballast resistors. Later experiments were made on a specially constructed discharge tube.

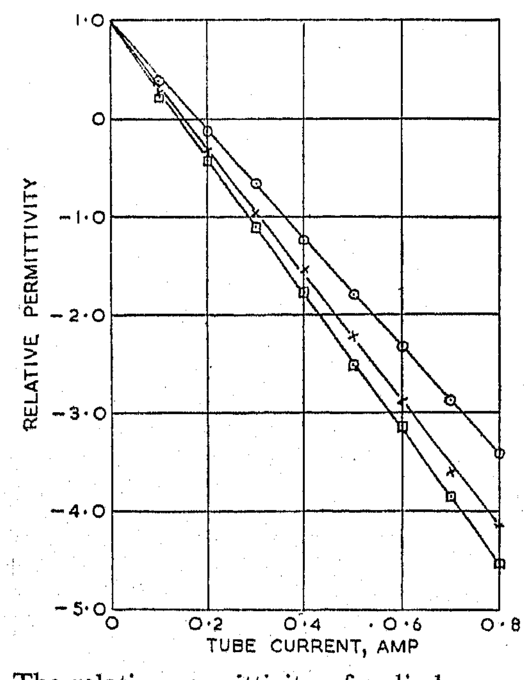


Fig. 2.—The relative permittivity of a discharge as a function of frequency and discharge current.

○ f_0 = 3 036 Mc/s.
× f_0 = 2 496 Mc/s.
⊡ f_0 = 2 114 Mc/s.

Fig. 2 shows the relative permittivity of the first discharge plotted as a function both of discharge current and impressed frequency. These results have been calculated ignoring the effects of the glass wall of the discharge tube. Assuming that

Dr. Foulds is with Marconi's Wireless Telegraph Co., Ltd.

the electron density is proportional to the discharge current, the results show the linear decrease of ϵ_1' with N, as indicated by eqns. (1) and (2). The relative slopes of the three lines of Fig. 2 do not bear out the relation of eqn. (3) and they show that the collisional frequency ν is not negligible compared with ω. The value of ν has been calculated by assuming that eqn. (1) holds at the three different frequencies. The results give $\nu = 2\cdot5 \times 10^{10}$ collisions/sec, and for a discharge current of 0·4 amp, $N = 7\cdot1 \times 10^{11}$ electrons/cm³. These values of ν and N agree satisfactorily with results quoted by Denno[15] and obtained from experiments in which a commercial discharge tube was used to scatter 3 cm-wavelength radiation.

A few experiments were carried out with an a.c. discharge, the tube being supplied from the 50 c/s mains. The results from one such experiment are shown in Fig. 3(*a*). The

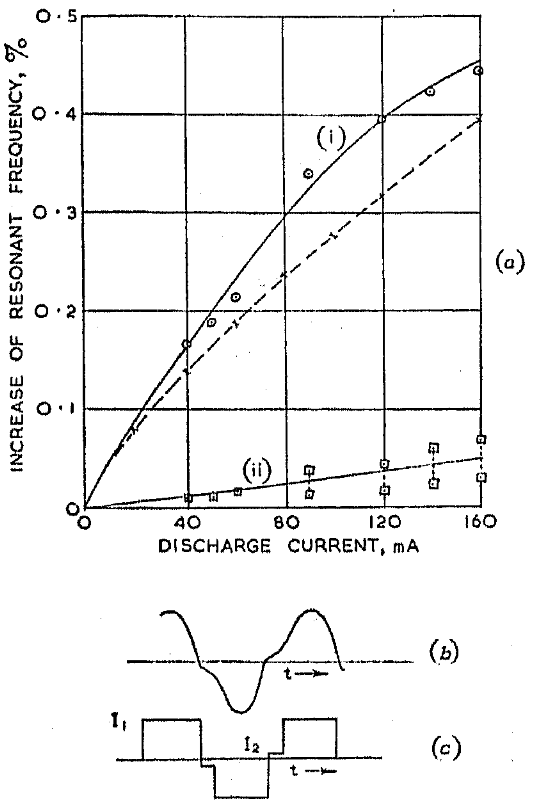


Fig. 3.—Cavity resonances and current waveforms for an a.c. discharge.
(*a*) Cavity resonances.
– – – D.C. discharge. ——— A.C. discharge.
(*b*) Typical current waveform of an a.c. discharge.
(*c*) Simplified current waveform of an a.c. discharge.

resonance curves of Fig. 3(*a*) are similar to those for the d.c. discharge, but the resonance curves of Fig. 3(*b*) become double-humped for values of I_a greater than 80 mA. The frequency difference between the two humps is shown by the dotted line between the two points at each current. It is shown in a qualitative manner that the resonance curves of Fig. 3(*b*) are probably caused by the step in the typical current waveform shown in that Figure and its simplified form of Fig. 3(*c*).

A rather interesting practical point arose in the experiments using the lighting tube when the discharge current was varied over the range 0–0·8 amp. It was always observed that the resonant frequency of the cavity for a given discharge current depended upon whether the current was increasing or decreasing. This hysteresis effect is shown very clearly in Fig. 4, and it is due primarily, if not completely, to the time lag between the temperature of the apparatus and the current in the discharge.

In order to repeat the experiments on a gaseous discharge in which the electron collisional frequency would be negligible compared with the impressed angular frequency, a special mercury-vapour discharge tube was constructed in which the pressure was that merely due to the saturation vapour pressure of the mercury. This tube is shown in Fig. 5. The discharge between

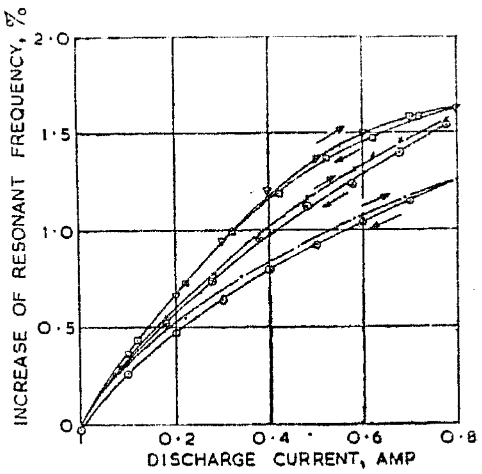


Fig. 4.—Variation of the resonant frequency of the cavity for increasing and decreasing currents.
△ and ⊡ $f_0 = 2114$ Mc/s.
× and ⊗ $f_0 = 2496$ Mc/s.
· and ⊙ $f_0 = 3036$ Mc/s.

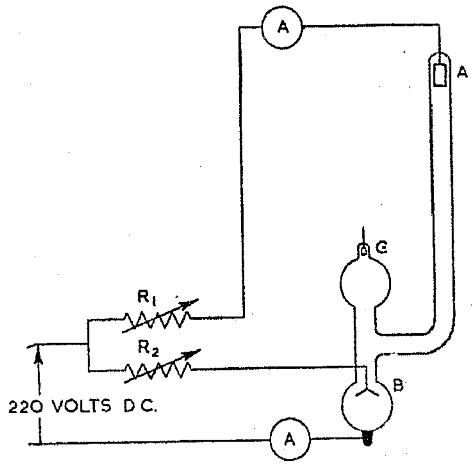


Fig. 5.—Circuit diagram of the low-pressure mercury-vapour tube.

the anode A and the mercury-pool cathode replaces the fluorescent lighting tube in the cavity, and the discharge between the anode B and the cathode is an auxiliary keep-alive system. The cathode pool was immersed in a thermostatically controlled water bath; this enabled the pressure within the tube to be controlled conveniently over the range 0·0007–0·0057 mm Hg. The temperature lag and the associated hysteresis mentioned previously were avoided by enclosing both the tube and the cavity in a hot-air jacket.

For discharge currents greater than 50 mA the resonance curves indicated by a galvanometer became markedly asymmetrical and damped; some typical examples are shown in Fig. 6. It is shown that this is due to fluctuations in the discharge and that it is associated with fluctuations in the light intensity emitted by the discharge. These fluctuations are the result of the interaction of two sets of striae. One set is almost regular and travels towards the cathode, and the other occurs apparently at random and travels towards the anode. Similar but more regular fluctuations have been fully examined by Donahue and Dieke.[21]

In subsequent experiments using this low-pressure mercury-vapour tube the instantaneous resonance curves were displayed on an oscillograph. This enabled the true heights of the resonance curves to be measured over the range of current in which the fluctuations occurred.

Fig. 7 shows the relative permittivity plotted as a function of discharge current, resonant frequency, and water-bath temperature. The results have been calculated taking into account the effect of the glass wall of the discharge tube. The slopes of the lines, for either water-bath temperature, again do not conform to the $1/\omega^2$ relation indicated by eqn. (3), and to explain this

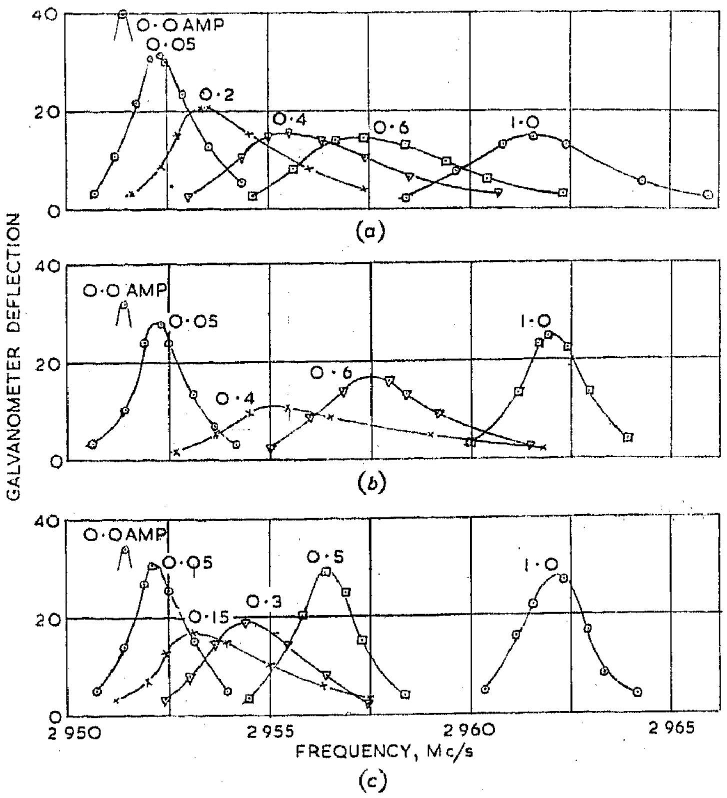


Fig. 6.—Resonance curves of the cavity as a function of discharge current and temperature of the water bath.
(a) $T = 20°$ C. (b) $T = 29°$ C. (c) $T = 36°$ C.

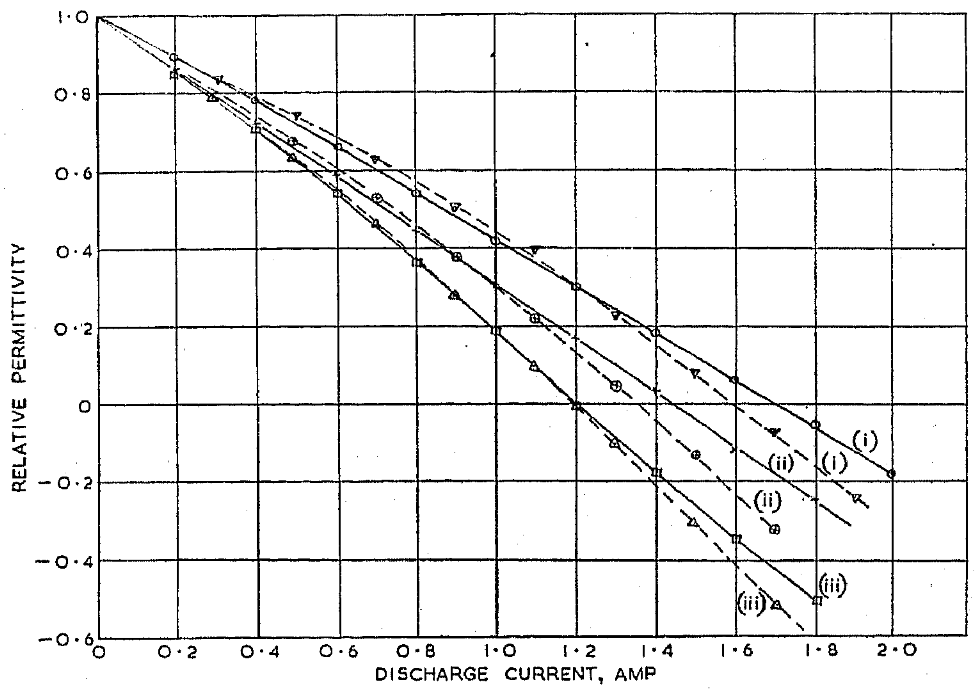


Fig. 7.—The relative permittivity of the discharge as a function of discharge current, water-bath temperature and frequency.
(i) $f_0 = 2946$ Mc/s, (ii) $f_0 = 2500$ Mc/s, (iii) $f_0 = 2119$ Mc/s.
——— $T = 20°$ C. - - - - $T_0 = 38{\cdot}6°$ C.

by stipulating a large collisional frequency requires ν to be $1{\cdot}5 \times 10^{10}$ collisions/sec. Such a large value is inconsistent with the small damping actually observed. Another discrepancy from the expected results is that the relative permittivity of the discharge does not depend very much upon the temperature of the water bath, but this can be explained by the hypothesis that the effective mercury free path of the electrons is limited by the physical size of the discharge tube.

The height of the resonance curve is proportional to the square of the Q-factor, and therefore the variation of the Q-factor with discharge current can be calculated from the curves of Fig. 8, which show the relation between the height of the resonance curve and discharge current. These curves suggest that the Q-factor decreases approximately uniformly for increases in the discharge current, but that superimposed upon this is a pronounced dip caused by some as yet unspecified phenomenon. It is shown in the appendix to the paper that ν is related to the regular decrease in the Q-factor by the equation

$$\nu = \frac{\omega}{2}\frac{f_0}{\delta f}\frac{1}{Q_0}\left(\frac{Q_0}{Q_L} - 1\right) \quad . \quad . \quad . \quad . \quad (4)$$

Ignoring the results at 3 000 Mc/s because the pronounced dip seriously affects the experimental results, the average value of ν deduced for the current of 1·8 amp is $0{\cdot}94 \times 10^8$ collisions/sec.

Corresponding to this value of ν, the Q-factor, and therefore the deflection at resonance, can be calculated for any discharge current directly from experimental curves of $\delta f \times \nu \times I_a$. The calculated relations between θ_L/θ_0 and I_a are shown by the full lines of Fig. 8, and the agreement is satisfactory in the regions

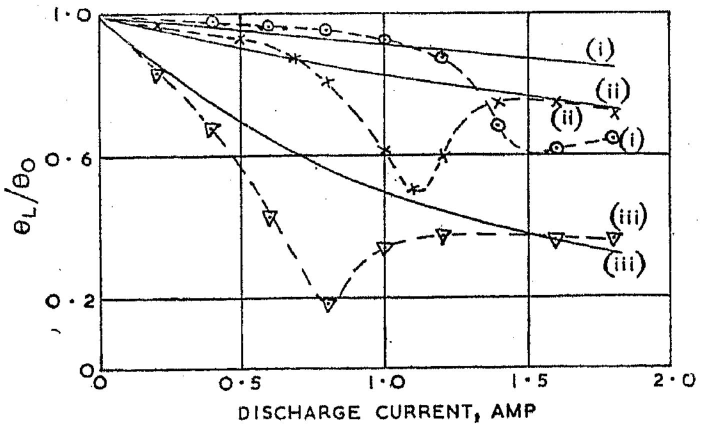


Fig. 8.—Curves of θ_L/θ_0 versus I_a.
$= 38{\cdot}6°$ C.
(i) $f_0 = 2946$ Mc/s, (ii) $f_0 = 2500$ Mc/s, (iii) $f_0 = 2119$ Mc/s.
——— Theoretical. - - - - Experimental.

not affected by the pronounced dip. This new value of ν also supports the suggestion that the mean free path is limited by the physical size of the discharge tube. It is shown qualitatively that the pronounced dip in the Q-factor may be caused by an electron resonance phenomenon of the kind envisaged by Tonks and Langmuir.[25] Other possible explanations are discussed and shown to be unlikely.

Under the action of the impressed r.f. electric field, electron resonance of the kind just mentioned would be associated with vigorous oscillations of those electrons within a certain annulus, and the electrons nearer the axis would therefore be partially shielded from the electric field. Under these conditions it is not possible to calculate the relative permittivity of the discharge by the analysis given in the appendix. This would account for the departure from the $1/\omega^2$ relation.

In the appendix are also derived the equations which enable the relative permittivity of loss-free and homogeneous discharge to be calculated from the change of the resonant frequency of an H_{011} resonant cavity. This analysis is rigorous and extends the results of Pincherle[26] to the case of a cylindrical waveguide system containing three concentric dielectrics. An approximate analysis based on the resonator action theorem of Maclean[30] shows how the relative permittivity of certain inhomogeneous discharges can be calculated from the measured change of the resonant frequency of the cavity. This theorem is also used to deduce the relation between the Q-factor of the cavity and the collisional frequency of the electrons.

Bibliography

(1) Eccles, W. H.: "On the Diurnal Variation of the Electric Waves occurring in Nature, and on the Propagation of Electric Waves round the Bend of the Earth," *Proceedings of the Royal Society*, 1912, **A87**, p. 79.

(2) Larmor, J.: "Why Wireless Electric Rays can bend round the Earth," *Philosophical Magazine*, 1924, **48**, p. 1025.

(3) MIMNO, H. R.: "The Physics of the Ionosphere," *Review of Modern Physics*, 1937, **9**, p. 1.
(4) VAN DER POL: Thesis, University of Utrecht, 1920.
(5) GUTTON, H., and CLEMENT, J.: "Dielectric Properties of Ionised Gases," *Comptes rendus hebdomadaires des séances de l'Academie des sciences*, 1927, **184**, p. 441, and "The Propagation of Electromagnetic Waves round the Earth," *ibid.*, 1927, **184**, p. 676.
(6) APPLETON, E. V., and CHILDS, E. C.: "On Some Radio-Frequency Properties of Ionized Air," *Philosophical Magazine*, 1930, **10**, p. 969.
(7) APPLETON, E. V., and CHAPMAN, F. W.: "The Collisional Frequency experienced by Vibrating Electrons in Ionized Air," *Proceedings of the Physical Society*, 1932, **44**, p. 246.
(8) BANNERJEE, S. S.: "On the Resonance Frequency of Oscillatory Circuits with Leaky Condenser and its Bearing on Measurement of the Dielectric Constant of Ionized Gas," *Philosophical Magazine*, 1934, **17**, p. 834.
(9) MITRA, S. K., and BANNERJEE, S. S.: "Dielectric Constant of Ionized Air," *Nature*, 1935, **136**, p. 512.
(10) IMAM, A., and KHASTGIR, S. R.: "Dielectric Constant of Ionized Gases," *Philosophical Magazine*, 1937, **23**, p. 858.
(11) GANGOPADHYAYA, S., and KHASTGIR, S. R.: "Dielectric Constant of Ionized Gases," *ibid.*, 1938, **25**, p. 883.
(12) KHASTGIR, S. R., and RAHMAN, S. M. F.: "The Dielectric Constant and Electrical Conductivity of Gases and Vapours Ionized by X-Rays at Ultra-High Radio Frequency," *ibid.*, 1940, **29**, p. 353.
(13) HORNER, F., TAYLOR, T. A., DUNSMUIR, R., LAMB, J., and WILLIS JACKSON: "Resonance Methods of Dielectric Measurement at Centimetre Wavelengths," *Journal I.E.E.*, 1946, **93**, Part III, p. 53.
(14) ADLER, F. P.: "Measurement of the Complex Conductivity of an Ionized Gas at Microwave Frequencies," *Journal of Applied Physics*, 1949, **20**, p. 1125.
(15) DENNO, S. N.: Ph.D. Thesis, University of Liverpool, 1951.
(16) COTTON, H.: "Electric Discharge Lamps" (Chapman and Hall, 1946).
(17) FRANCIS, V. J.: "Fundamental of Discharge Tube Circuits" (Methuen and Co., Ltd., 1948).
(18) LANGMUIR, I.: "Positive Ion Currents in the Positive Column of the Mercury Arc," *General Electric Review*, 1923, **26**, p. 731.
(19) KILLIAN, T. J.: "The Uniform Positive Column of an Electric Discharge in Mercury Vapour," *Physical Review*, 1930, **35**, p. 1238.
(20) CULLEN, A. L., and FOULDS, K. W. H.: "Resonant Cavity Measurements on a D.C. Discharge," *Nature*, 1952, **169**, p. 236.
(21) DONAHUE, T., and DIEKE, G. H.: "Oscillatory Phenomena in Direct Current Glow Discharges," *Physical Review*, 1951, **81**, p. 248.
(22) DENNO, S. N., PRIME, H. A., and CRAGGS, J. D.: "The Scattering of 3 cm Radiation by Ionized Gases," *Proceedings of the Physical Society*, 1950, **63B**, p. 726.
(23) ROMMEL, D.: "Radio Reflections from a Column of Ionized Gas," *Nature*, 1951, **167**, p. 243.
(24) KAISER, T. R., and CLOSS, R. L.: "Theory of Radio Reflections from Meteor Trails," *Philosophical Magazine*, 1952, **43**, p. 1.
(25) TONKS, L., and LANGMUIR, I.: "Oscillations in Ionized Gases," *Physical Review*, 1929, **33**, p. 195.
(26) PINCHERLE, L.: "Electromagnetic Waves in Metal Tubes filled longitudinally with Two Dielectrics," *ibid.*, 1944, **66**, p. 118.
(27) COLLIE, C. H., HASTED, J. B., and RITSON, D. M.: "The Cavity Resonator Method of Measuring the Dielectric Constant of Polar Liquids in the Centimetre Band," *Proceedings of the Physical Society*, 1948, **60**, p. 71.
(28) ROSEN, P.: "The Propagation of Electromagnetic Waves in a Tube containing a Coaxial D.C. Discharge," *Journal of Applied Physics*, 1949, **20**, p. 868.
(29) British Association Mathematical Tables, Volume 6 (Cambridge University Press, 1937).
(30) MACLEAN, W. R.: "The Resonator Action Theorem," *Quarterly of Applied Mathematics*, 1945, **2**, p. 329.
(31) BARLOW, H. E. M., and CULLEN, A. L.: "Microwave Measurements" (Constable and Company Ltd., 1950).

THE STABILITY AND TIME RESPONSE OF FAST OPERATING CLOSED-LOOP PULSED RADAR CIRCUITS

621.396.96-526 : 621.3.016.35 Monograph No. 121 R

D. McDONNELL and W. R. PERKINS, B.Sc.

(DIGEST *of a paper published in March*, 1955, *as an* INSTITUTION MONOGRAPH *and to be republished in Part C of the* PROCEEDINGS.)

The type of sampling system analysed in the paper is shown in Fig. 1. The system is split into three parts as follows:

(*a*) *Sampler.*—This unit samples the difference between the input and output and maintains the sampled value until the next sampling time. (The intervals between the sampling times are equal and are normalized to unity throughout the paper.)

(*b*) *Discontinuous Shaping Unit.*—In this unit the output from the "sampler" is modified by any linear combination of "summing" and "differencing." It is also possible for the information

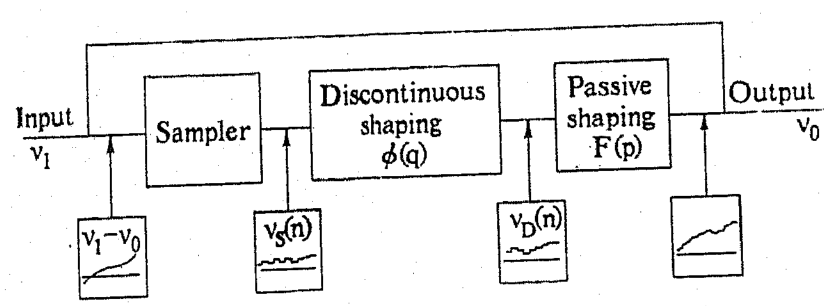


Fig. 1.—Block diagram of the sampling servo-system.
Waveforms are only symbolic.

Mr. McDonnell and Mr. Perkins are with Vickers-Armstrong, Ltd.

to be delayed by an integral number of sampling periods by the unit.

(c) *Passive Shaping Unit.*—This is any linear network consisting of L, C and R elements which has a step response $f(t)$ whose Laplace transform is $F(p)$.

The output of the system at time $t = n + \delta t$ is found to be given by

$$v_0(n + \delta t) = \frac{1}{2\pi i} \int_c \frac{\theta(q) \sum^r B_r \varepsilon^{\alpha_r \delta t}/(q - \varepsilon^{\alpha_r}) \mathscr{D}[v_i(n)] q^{n-1}}{\sum^r B_r/(q - \varepsilon^{\alpha_r})} dq$$

where the contour c is a large circle around the origin, $\theta(q)$, B_r and α_r are parameters describing the kind of units used within the sampling servo system, $\theta(q)$ is a rational polynomial, B_r and α_r are constants and

$$\mathscr{D}[v_i(n)] = \sum_{n=0}^{n=\infty} q^{-n} v_i(n)$$

Nyquist plots can be made, when it is possible to expand $\theta(q)$ about the point $q = 1$ in the form

$$\theta(q) = \sum_{n=0}^{n=r_1} A(n)(q-1)^n + \sum_{n=0}^{n=r_2} B(n) q^n/(q-1)^n$$

The paper includes examples using simple discontinuous shaping units which are typical of range measuring units and a.g.c. networks. The system is discussed from a stochastic standpoint in the appendix, and an expression for the spectral density of the output is obtained. A further example is included to illustrate the significance of this result. Some of the work included in the paper has been discussed by other writers on several occasions, but the main features in this field which the authors consider as new are as follows:

(a) The generalized discontinuous shaping unit and its complete inclusion in the mathematics.

(b) The expression for the output *between* the sampling times for the general case.

(c) The conditions under which Nyquist plots can be made.

(d) A comprehensive study of the time response to a stochastic process with its accompanying results.

The paper is almost entirely mathematical and should prove of great interest to servo-mechanism specialists.

REFERENCES

(1) VAN VOORHIS, N. S.: "Microwave Receivers" (McGraw-Hill Book Co.), M.I.T., Vol. 23, p. 407.

(2) WIENER, N.: "Generalized Harmonic Analysis," *Acta Mathematica*, 1930, **55**, p. 203.

(3) "Automatic and Manual Control" (Butterworth, 1952), p. 377.

(4) COOPER, S. R.: "A Study of a Second-order Sampling Servo," *Electronic Engineering*, 1953, **25**, p. 342.

(5) MCCOLL, LEROY A.: "Fundamental Theory of Servo Mechanisms" (D. Van Nostrand, 1945).

(6) MILNE, W. E.: "Numerical Calculus" (Princeton University Press, 1949).

(7) JAMES, H. M., NICHOLS, N. B., and PHILLIPS, R. S.: "Theory of Servo Mechanisms" (McGraw-Hill Book Co., 1947), M.I.T., Vol. 25.

(8) MCLACHLAN, N. W.: "Complex Variable and Operational Calculus" (Cambridge University Press, 1947).

(9) PHILLIPS, E. G.: "Functions of a Complex Variable" (Oliver and Boyd Mathematical Texts, 1940).

(10) TITCHMARSH, E. C.: "Theory of Functions" (Oxford University Press, 1932).

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING), JULY 1955

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part. *Example.*—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 S, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.